Modeling Function-Valued Stochastic Processes, With Applications to Fertility Dynamics

Kehui Chen[1], Pedro Delicado[2] and Hans-Georg Müller[3]

[1]Dept. of Statistics, University of Pittsburgh, Pittsburgh, USA
[2]Dept. d'Estadística i Inv. Op., Universitat Politècnica de Catalunya, Barcelona, Spain
[3]Department of Statistics, University of California, Davis, USA

October 27, 2015

ABSTRACT

We introduce a simple and interpretable model for functional data analysis for situations where the observations at each location are functional rather than scalar. This new approach is based on a tensor product representation of the function-valued process and utilizes eigenfunctions of marginal kernels. The resulting marginal principal components and product principal components are shown to provide optimal representations in a well-defined sense. Given a sample of independent realizations of the underlying function-valued stochastic process, we propose straightforward fitting methods to obtain the components of this model and to establish asymptotic consistency and rates of convergence for the proposed estimates. The methods are illustrated by modeling the dynamics of annual fertility profile functions for 17 countries. This analysis demonstrates that the proposed approach leads to insightful interpretations of the model components and interesting conclusions.

KEY WORDS: Asymptotics, demography, functional data analysis, functional principal components, product principal component analysis, tensor product representation.

This research was supported by NSF grants DMS-1104426, DMS-1228369, DMS-1407852, by the Spanish Ministry of Education and Science, and FEDER grant MTM2010-14887. The main part of this work was done when Pedro Delicado was visiting UC Davis with the financial support of the Spanish Government (Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i).U.S.

1. INTRODUCTION

In various applications one encounters stochastic processes and random fields that are defined on temporal, spatial or other domains and take values in a function space, assumed to be the space of square integrable functions L^2. More specifically, for $\mathcal{S} \subset \mathbb{R}^{d_1}$ and $\mathcal{T} \subset \mathbb{R}^{d_2}$, we consider the stochastic process $X : \mathcal{T} \to L^2(\mathcal{S})$ and denote its value at *time* $t \in \mathcal{T}$ by $X(\cdot, t)$, a square integrable random function with argument $s \in \mathcal{S}$. A key feature of our approach is that we consider the case where one has n independent observations of the functional stochastic process.

A specific example that we will discuss in detail below (see Section 5) is that of female fertility profile functions $X(\cdot, t)$, available annually (t = year) for $n = 17$ countries, with age as argument s. The starting point is the *Age-Specific Fertility Rate* (ASFR) $X(s, t)$ for a specific country, defined as

$$X(s,t) = \text{ASFR}(s,t) = \frac{\text{Births during the year } t \text{ given by women aged } s}{\text{Person-years lived during the year } t \text{ by women aged } s}. \qquad (1)$$

Figure 1 illustrates the ASFR data for the U.S. from 1951 to 2006. The left panel shows ASFR$(\cdot, t)$ for t = 1960, 1980 and 2000. The image plot representing ASFR(s, t) for all possible values of s and t in the right panel provides a visualization of the dynamics of fertility in the U.S. over the whole period.

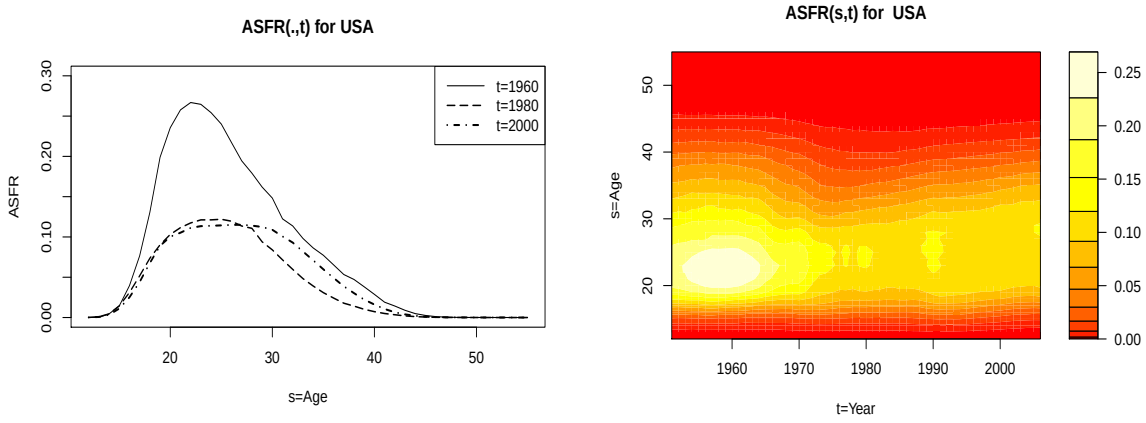

Figure 1: Age Specific Fertility Rate for the U.S. *Left:* Profiles for three calendar years. *Right:* Image plot representing ASFR(s, t) for all possible values of s and t.

For data structures where one observes only one realization of a function-valued process, related modeling approaches have been discussed previously (Delicado *et al.* 2010; Nerini *et al.* 2010; Gromenko *et al.* 2012, 2013; Huang *et al.* 2009). Similarly, Hyndman and Ullah (2007) and Hyndman and Shang (2009) considered functional time series in a setting where only one realization is observed. In related applications such as mortality analysis, the decomposition into age and year has been studied by Eilers and Marx (2003); Currie *et al.* (2004, 2006); Eilers *et al.* (2006), using P-splines. The case where i.i.d. samples are available for random fields has been much less studied. Multilevel functional models and functional mixed effects models have been investigated by Morris and Carroll (2006), Crainiceanu *et al.* (2009), Greven *et al.* (2010), and Yuan *et al.* (2014), among others, while Chen and Müller (2012) developed a "double functional principal component" method and studied its asymptotic properties.

Our approach applies to general dimensions of both the domain of the underlying random process, with argument t, as well as of the domain of the observed functions, with argument s, while we emphasize the case of function-valued observations for stochastic processes on a one-dimensional time domain. This is the most common case and it often allows for particularly meaningful interpretations. Consider processes $X(s,t)$ with mean $\mu(s,t) = E(X(s,t))$ for all $s \in \mathcal{S} \subseteq \mathbb{R}^{d_1}$ and all $t \in \mathcal{T} \subseteq \mathbb{R}^{d_2}$, and covariance function

$$C((s,t),(u,v)) = E(X(s,t)X(u,v)) - \mu(s,t)\mu(u,v) = E(X^c(s,t)X^c(u,v)), \tag{2}$$

where here and in the following we denote the centered processes by X^c.

A well-established tool of Functional Data Analysis (FDA) is Functional Principal Component Analysis (FPCA) (Ramsay and Silverman 2005) of the random process $X(s,t)$, which is based on the Karhunen-Loève expansion

$$X(s,t) = \mu(s,t) + \sum_{r=1}^{\infty} Z_r \gamma_r(s,t), \quad s \in \mathcal{S},\, t \in \mathcal{T}. \tag{3}$$

Here $\{\gamma_r : r \geq 1\}$ is an orthonormal basis of $L^2(\mathcal{S} \times \mathcal{T})$ that consists of the eigenfunctions of the covariance operator of X, and $\{Z_r = \int \gamma_r(s,t)X^c(s,t)dsdt : r \geq 1\}$ are the (random)

coefficients. This expansion has the optimality property that the first K terms form the K-dimensional representation of $X(s,t)$ with the smallest unexplained variance.

A downside of the two- or higher-dimensional Karhunen-Loève representation (3) is that it allows only for a joint symmetric treatment of the arguments and therefore is not suitable for analyzing the separate (possibly asymmetric) effects of s and t. An additional technical drawback is that an empirical version of (3) requires the estimation of the covariance function C in (2) that depends on dimension $2(d_1 + d_2)$, and for the case of sparse designs, this then requires to perform non-parametric regression depending on at least four variables, with associated slow computing, curse of dimensionality and loss of asymptotic efficiency. Finally, Karhunen-Loève expansions for functional data depending on more than one argument are non-standard and suitable software is hard to obtain.

Aiming to address these difficulties and with a view towards interpretability and simplicity of modeling, we propose in this paper the following representation,

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \xi_j(t)\psi_j(s) = \mu(s,t) + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \chi_{jk}\phi_{jk}(t)\psi_j(s), \tag{4}$$

where $\{\psi_j : j \geq 1\}$ are the eigenfunctions of the operator in $L^2(\mathcal{S})$ with kernel

$$G_{\mathcal{S}}(s,u) = \int_{\mathcal{T}} C((s,t),(u,t))dt, \tag{5}$$

while $\{\xi_j(t) : j \geq 1\}$ are the (random) coefficients of the expansion of the centered processes $X^c(\cdot,t)$ in $\psi_j(s)$, and $\xi_j(t) = \sum_{k=1}^{\infty} \chi_{jk}\phi_{jk}(t)$ is the Karhunen-Loève expansion of the random functions $\xi_j(t)$ in $L^2(\mathcal{T})$ with eigenfunctions ϕ_{jk} and FPCs χ_{jk}.

We refer to $G_{\mathcal{S}}$ as *the marginal covariance function*, and to (4) as the *marginal Karhunen-Loève representation* of X that leads to the *marginal FPCA* and note that the product basis functions $\phi_{jk}(t)\psi_j(s)$ are orthogonal to each other. Hence the scores χ_{jk} can be optimally estimated by the inner product of X^c with the corresponding basis. Also, for each $j \geq 1$, we have $E\chi_{jk}\chi_{jk'} = 0$ for $k \neq k'$. In related settings, marginal covariance functions very recently have also been utilized by other researchers (Park and Staicu 2015; Aston *et al.* 2015). In Theorem 1 below we establish the optimality of the marginal

eigenfunctions ψ_j under a well-defined criterion and show in Theorem 2 that the finite expansion based on the marginal FPCA approach nearly minimizes the variance among all representations of the same form.

When using the representation (4), the effects of the two arguments s and t can be analyzed separately, which we will show in greater detail below in Sections 2 and 5. We also note that the estimation of the marginal representation (4) requires only to estimate covariance functions that depend on $2d_1$ or $2d_2$ real arguments. In particular, when $d_1 = d_2 = 1$, only two-dimensional surfaces need to be estimated and marginal FPCA can be easily implemented using standard functional data analysis packages.

Motivated by a common principal component perspective, we also introduce a simplified version of (4), the *product FPCA*,

$$X(s,t) = \mu(s,t) + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \chi_{jk}\phi_k(t)\psi_j(s), \tag{6}$$

where the ϕ_k, $k \geq 1$, are the eigenfunctions of the marginal kernel $G_{\mathcal{T}}(s,u)$, analogous to $G_{\mathcal{S}}(t,v)$, with supporting theory provided by Theorem 4 and Theorem 5.

Sections 2 and 3 provide further details on model and estimation. Theoretical considerations are in Section 4. In Section 5, we compare the performance of the proposed marginal FPCA, product FPCA and the conventional two-dimensional FPCA in the context of an analysis of the fertility data. Simulation results are described in Section 6 and conclusions can be found in Section 7. Detailed proofs, additional materials and the analysis of an additional human mortality data example have been relegated to the Online Supplement.

2. MARGINAL FPCA

2.1. Modeling

Consider the standard inner product, $\langle f, g\rangle = \int_{\mathcal{S}}\int_{T} f(s,t)g(s,t)dtds$ in the separable Hilbert space $L^2(\mathcal{S}\times\mathcal{T})$ and the corresponding norm $\|\cdot\|$. In the following, X is in $L^2(\mathcal{S}\times\mathcal{T})$ with mean $\mu(s,t)$. Using the covariance function $C((s,t),(u,v))$ as kernel for

the Hilbert-Schmidt covariance operator $\Gamma(f)(s,t) = \int_{\mathcal{S}} \int_{\mathcal{T}} C((s,t),(u,v)) f(u,v) dv\, du$ with orthonormal eigenfunctions γ_r, $r \geq 1$, and eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ then leads to the Karhunen-Loève representation of X in (3), where $E(Z_r) = 0$ and $\text{cov}(Z_r, Z_l) = \lambda_r \delta_{rl}$, with $\delta_{rl} = 1$ for $r = l$ and $= 0$ otherwise; see Horváth and Kokoszka (2012) and Cuevas (2013).

Since the marginal kernel $G_{\mathcal{S}}(s,u)$ as defined in (5) is a continuous symmetric positive definite function (see Lemma 1 in Online Supplement A), denoting its eigenvalues and eigenfunctions by τ_j, ψ_j, $j \geq 1$, respectively, the following representation for X emerges,

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \xi_j(t) \psi_j(s), \tag{7}$$

where $\xi_j(t) = \langle X(\cdot,t) - \mu(\cdot,t), \psi_j \rangle_{\mathcal{S}}$, $j \geq 1$, is a sequence of random functions in $L^2(\mathcal{T})$ with $E(\xi_j(t)) = 0$ for $t \in \mathcal{T}$, and $E(\langle \xi_j, \xi_k \rangle_{\mathcal{T}}) = \tau_j \delta_{jk}$ (see Lemma 2 in Online Supplement A). Theorem 1 in Section 4 shows that the above representation has an optimality property.

The marginal Karhunen-Loève representation (7) provides new functional data, the score functions $\xi_j(t)$, which are random functions that depend on only one argument. For each $j \geq 1$, the ξ_j have their own covariance functions $\Theta_j(t,v) = E(\xi_j(t)\xi_j(v))$, $t, v \in \mathcal{T}$, $j \geq 1$, with eigencomponents (eigenvalues/eigenfunctions) $\{\eta_{jk}, \phi_{jk}(t) : k \geq 1\}$. The continuity of the covariance function C implies that the $\Theta_j(t,v)$ are also continuous functions. The random functions $\xi_j(t)$ then admit their own Karhunen-Loève expansions,

$$\xi_j(t) = \sum_{k=1}^{\infty} \chi_{jk} \phi_{jk}(t), \quad j \geq 1, \tag{8}$$

with $E(\chi_{jk}) = 0$ and $E(\chi_{jk}\chi_{jr}) = \eta_{jk}\delta_{kr}$. From (7) and (8) we obtain the representation for $X(s,t)$ in (4), $X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \chi_{jk} \phi_{jk}(t) \psi_j(s)$. As already mentioned, this expansion does not coincide with the standard Karhunen-Loève expansion of X and it is not guaranteed that χ_{jk} and χ_{lr} are uncorrelated for $j \neq l$. But the product functions $\phi_{jk}(t)\psi_j(s)$ remain orthonormal in the sense that $\int_{\mathcal{S},\mathcal{T}} \phi_{jk}(t)\psi_j(s)\phi_{lh}(t)\psi_l(s) ds dt = \delta_{jk,lh}$, where $\delta_{jk,lh} = 1$ when $j = l$ and $k = h$; zero otherwise.

2.2 Estimating Procedures

Time- or space-indexed functional data consist of a sample of n independent subjects or

units. For the i-th subject, $i = 1, \ldots, n$, random functions $X_i(\cdot, t)$ are recorded at a series of time points t_{im}, $m = 1, \ldots, M_i$. Ordinarily, these functions are not continuously observed, but instead are observed at a grid of functional design points s_l, $l = 1, \ldots, L$. In this paper we focus on the case where the grid of s is dense, regular and the same across all subjects. The case of sparse designs in s will be discussed in Section 7. Our proposed marginal FPCA procedure consists of three main steps:

Step 1. Center the data to obtain $\hat{X}_i^c(s,t) = X_i(s,t) - \hat{\mu}(s,t)$. Obtain an estimator of $\mu(s,t)$ by pooling all the data together. If the recording points t are densely and regularly spaced, i.e., $t_{im} = t_m$, an *empirical estimator* by averaging over n subjects and interpolating between design points can be used. This scheme is also applicable to dense irregular designs by adding a pre-smoothing step and sampling smoothed functions at a dense regular grid. Alternatively, one can recover the mean function μ by smoothing the pooled data (Yao *et al.* 2005), for example with a local linear smoother, obtaining a *smoothing estimator* $\hat{\mu}(s,t) = \hat{a}_0$, where

$$(\hat{a}_0, \hat{a}_1, \hat{a}_2) = \arg\min \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{M_i} \sum_{l=1}^{L_{im}} \{[X_i(t_{im}, s_{iml}) - a_0 - a_1(s_{iml} - s) - a_2(t_{im} - t)]^2 \times K_{h_s}(s_{iml} - s) K_{h_t}(t_{im} - t)\}. \tag{9}$$

Step 2. Use the centered data $\hat{X}_i^c(s,t)$ from Step 1 to obtain estimates of the marginal covariance function $G_{\mathcal{S}}(s,u)$ as defined in (5), its eigenfunctions $\psi_j(s)$ and the corresponding functional principal component (FPC) score functions $\xi_{i,j}(t)$. For this, we pool the data $\{\hat{X}_i^c(\cdot, t_{im}),\ i = 1, \ldots, n,\ m = 1, \ldots, M_i\}$ and obtain estimates

$$\hat{G}_{\mathcal{S}}(s_j, s_l) = \frac{|\mathcal{T}|}{\sum_{i=1}^{n} M_i} \sum_{i=1}^{n} \sum_{m=1}^{M_i} \hat{X}_i^c(s_j, t_{im}) \hat{X}_i^c(s_l, t_{im}), \tag{10}$$

where $1 \le j \le l \le L$ and $|\mathcal{T}|$ is the Lebesgue measure of $\mathcal{T}$, followed by interpolating between grid points to obtain $\hat{G}_{\mathcal{S}}(s,u)$ for $(s,u) \in \mathcal{S} \times \mathcal{S}$. One then obtains the eigenfunctions $\hat{\psi}_j$ and eigenvalues $\hat{\tau}_j$ by standard methods (Yao *et al.* 2005) as implemented in the PACE package (http://www.stat.ucdavis.edu/PACE) or as in Kneip and Utikal (2001), and the

FPC function estimates $\hat{\xi}_{i,j}(t)$ by interpolating numerical approximations of the integrals $\hat{\xi}_{i,j}(t_{im}) = \int \hat{X}_i^c(s,t_{im})\hat{\psi}_j(s)ds$. Theorem 3 shows that $\hat{G}_{\mathcal{S}}$ in (10) and $\hat{\psi}_j$ are consistent estimates of the marginal covariance function $G_{\mathcal{S}}$ and its eigenfunctions and that estimates $\{\hat{\xi}_{i,j}(t),\ i = 1,\ldots,n,\}$ converge uniformly to the target processes $\{\xi_{i,j}(t), j \geq 1\}$.

Step 3. This is a standard FPCA of one-dimensional processes $\{\xi_{i,j}(t), j \geq 1\}$, where for each fixed j, one obtains estimates for the FPCs χ_{jk} and eigenfunctions $\{\phi_{jk}(t) : k \geq 1\}$; see for example Ramsay and Silverman (2005); Kneip and Utikal (2001) for designs that are dense in t and Yao *et al.* (2005) for designs that are sparse in t.

After selecting appropriate numbers of included components P and K_j, $j = 1,\ldots,P$, one obtains the overall representation

$$\hat{X}_i(s,t) = \hat{\mu}(s,t) + \sum_{j=1}^{P} \hat{\xi}_{i,j}(t)\hat{\psi}_j(s) = \hat{\mu}(s,t) + \sum_{j=1}^{P}\sum_{k=1}^{K_j} \hat{\chi}_{i,jk}\hat{\phi}_{jk}(t)\hat{\psi}_j(s). \tag{11}$$

The included number of components P can be selected via a fraction of variance explained (FVE) criterion, finding the smallest P such that $\sum_{j=1}^{P}\hat{\tau}_j / \sum_{j=1}^{M}\hat{\tau}_j \geq 1-p$, where M is large and we choose $p = 0.15$ in our application. The number of included components K_j can be determined by a second application of the FVE criterion, where the variance V_{jk} explained by each term (j,k) is defined as

$$V_{jk} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{\chi}_{i,jk}^2}{\frac{1}{n}\sum_{i=1}^{n}||X(s,t) - \hat{\mu}(s,t)||_{\mathcal{S}\times\mathcal{T}}^2}. \tag{12}$$

Note that V_{jk} does not depend on the choice of P in the first step, since it is the fraction of total variance explained. Here total variance explained, $\sum_{k=1}^{K_j}\sum_{j=1}^{P} E(\chi_{jk}^2)$, cannot exceed the variance explained in the first step, $\sum_{j=1}^{P}\tau_j$.

We will illustrate these procedures in Section 5. Since the functions $\psi_j(s) \times \phi_{jk}(t)$ are orthogonal, the unexplained variance, $E\|X^c\|^2 - \sum_{j=1}^{P}\sum_{k=1}^{K_j} E(\chi_{jk}^2)$, and the reconstruction loss, $E\left(\int_{\mathcal{S},\mathcal{T}}\{X^c(s,t) - \sum_{j=1}^{P}\sum_{k=1}^{K_j}\langle X^c, \phi_{jk}\psi_j\rangle\phi_{jk}(t)\psi_j(s)\}^2 dsdt\right)$, are equivalent.

3. PRODUCT FPCA

In this section we discuss a simplified version of the marginal Karhunen-Loève representation (4). A simplifying assumption is that the eigenfunctions ϕ_{jk} in the Karhunen-Loève

expansion of $\xi_j(t)$ in (4) do not depend on j. This assumption leads to a more compact representation of X as given in (6), $X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \chi_{jk}\phi_k(t)\psi_j(s)$.

To study the properties of this specific product representation, we consider product representations with general orthogonal basis $X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \chi_{jk}f_k(t)g_j(s)$, where $\chi_{jk} = \langle X^c, f_k g_j\rangle$. For such general representations, the assumption

$$\text{cov}(\chi_{jk}, \chi_{jl}) = 0 \text{ for } k \neq l, \quad \text{and} \quad \text{cov}(\chi_{jk}, \chi_{hk}) = 0 \text{ for } j \neq h \tag{13}$$

implies that the covariance kernel induced by $\xi_j(t) = \langle X^c(t,\cdot), g_j\rangle_{\mathcal{S}}$ has common eigenfunctions $\{f_k(t), k \geq 1\}$, not depending on j, and the covariance kernel induced by $\xi_k(s) = \langle X^c(\cdot,s), f_k\rangle_{\mathcal{T}}$ has common eigenfunctions $\{g_j(s), j \geq 1\}$, not depending on k. Therefore we refer to (13) as the *common principal component assumption*. We prove in Theorem 4 below that if there exists bases $\{g_j(s), j \geq 1\}$ and $\{f_k(t), k \geq 1\}$ such that (13) is satisfied, then $g_j \equiv \psi_j$ and $f_k \equiv \phi_k$, the eigenfunctions of the marginal covariance $G_{\mathcal{S}}(s,u)$ and $G_{\mathcal{T}}(t,v)$, respectively, where $G_{\mathcal{T}}(t,v)$ is defined as

$$G_{\mathcal{T}}(t,v) = \int_{\mathcal{S}} C((s,t),(s,v))\, ds, \quad \text{with } t,v \in \mathcal{T}. \tag{14}$$

Even without invoking (13), in Theorem 5 we show that the finite expansion based on the marginal eigenfunctions ϕ_k and ψ_j yields a near-optimal solution in terms of minimizing the unexplained variance among all possible product expansions. This result provides additional theoretical support for the use of *product FPCA* based on the marginal kernels $G_{\mathcal{S}}$ and $G_{\mathcal{T}}$ under fairly general situations. While the product functions $\phi_k(t)\psi_j(s)$ are orthonormal, without addtional conditions, the scores χ_{jk} in general will not be uncorrelated. Product FPCA (6) is well suited for situations where the two arguments of $X(s,t)$ play symmetric roles. This simplified model retains substantial flexibility, as we will demonstrate in the application to fertility data (see Online Supplement C).

The estimation procedures for this model are analogous to those described in the previous section. This also applies to the theoretical analysis of these estimates and their asymptotic properties. A straightforward approach to estimate the eigenfunctions appearing in (6) is to apply the estimation procedure described in Section 2.2 twice, first following

the description there to obtain estimates of $G_{\mathcal{S}}$ and ψ_j and then changing the roles of the two arguments in a second step to obtain estimates of $G_{\mathcal{T}}$ and ϕ_k.

4. THEORETICAL PROPERTIES

Detailed proofs of the results in this section are in Online Supplement A. We show that the optimal finite-dimensional approximation property of FPCA extends to the proposed marginal FPCA under well defined criteria. Theorem 1 establishes the optimality of the basis functions ψ_j, i.e. the eigenfunctions in (4) derived from the marginal covariance in (5). Theorem 2 shows the near optimality of the marginal representation (4), based on the eigenfunctions ϕ_{jk} and ψ_j, in terms of minimizing the unexplained variance among all functional expansions of the same form.

Theorem 1. *For each $P \geq 1$ for which $\tau_P > 0$, the functions $g_1, \ldots, g_P$ in $L^2(\mathcal{S})$ that provide the best finite-dimensional approximations in the sense of minimizing*

$$E\left(\int_{\mathcal{T}} \|X^c(\cdot,t) - \sum_{j=1}^{P} \langle X^c(\cdot,t), g_j\rangle_{\mathcal{S}} g_j\|_{\mathcal{S}}^2 dt\right)$$

are $g_j = \psi_j$, $j = 1, \ldots, P$, i.e., the eigenfunctions of $G_{\mathcal{S}}$. The minimizing value is $\sum_{j=P+1}^{\infty} \tau_j$.

Theorem 2. *For $P \geq 1$ and $K_j \geq 1$, consider the following loss minimization*

$$\min_{f_{jk}, g_j} E\left(\int_{\mathcal{S},\mathcal{T}} \{X^c(s,t) - \sum_{j=1}^{P}\sum_{k=1}^{K_j} \langle X^c, f_{jk}g_j\rangle f_{jk}(t)g_j(s)\}^2 dsdt\right),$$

with minimizing value Q^*, where the $g_j(s)$, $j \geq 1$, are orthogonal and for each j, the $f_{jk}(t)$, $k \geq 1$ are orthogonal. The marginal eigenfunctions $\psi_j(s)$, and $\phi_{jk}(t)$ achieve good approximation in the sense that

$$E\left(\int_{\mathcal{S},\mathcal{T}} \{X^c(s,t) - \sum_{j=1}^{P}\sum_{k=1}^{K_j} \langle X^c, \phi_{jk}\psi_j\rangle \phi_{jk}(t)\psi_j(s)\}^2 dsdt\right) < Q^* + aE\|X^c\|^2,$$

where $a = \max_{1\leq j\leq P} a_j$, with $(1 - a_j)$ denoting the fraction of variance explained by K_j terms for each process $\xi_j(t) = \langle X^c(\cdot,t), \psi_j\rangle_{\mathcal{S}}$.

In the following, $\|G_{\mathcal{S}}(s,u)\|_{\mathcal{S}} = \{\int_{\mathcal{S}}\int_{\mathcal{S}}(G_{\mathcal{S}}(s,u))^2 dsdu\}^{1/2}$ is the Hilbert-Schmidt norm and $a \asymp b$ denotes that a and b are of the same order asymptotically. For the consistency of marginal FPCA (4) it is important that the covariance kernel $G_{\mathcal{S}}$ and its eigenfunctions ψ_j and eigenvalues τ_j can be consistently estimated from the data. Uniform convergence of the empirical working processes $\{\hat{\xi}_{i,j}(t_{im}), 1 \le i \le n, 1 \le m \le M_i\}$ to the target processes $\{\xi_{i,j}(t), t \in \mathcal{T}\}$ then guarantees the consistency of the estimates of the eigenfunctions ϕ_{jk} and the eigenvalues η_{jk} (Yao and Lee 2006).

The following assumptions are needed to establish these results. We use $0 < B < \infty$ as a generic constant that can take different values at different places.

(A.1) $\sup_{s,t}|\mu(s,t)| < B$ and $\sup_s|\psi_j(s)| < B$ for all $1 \le j \le P$.

(A.2) $E\sup_{s,t}|X(s,t)| < B$ and $\sup_{s,t}E|X(s,t)|^4 < B$.

(A.3) $\sup_{(s,u)\in\mathcal{S}^2,(t_1,t_2)\in\mathcal{T}^2}|C((s,t_1),(u,t_1)) - C((s,t_2),(u,t_2))| < B|t_1 - t_2|$

(A.4) $\sup_{(s_1,u_1,s_2,u_2)\in\mathcal{S}^4}|G_{\mathcal{S}}(s_1,u_1) - G_{\mathcal{S}}(s_2,u_2)| < B(|s_1 - s_2| + |u_1 - u_2|)$.

(A.5) For all $1 \le j \le P$, $\delta_j > 0$, where $\delta_j = \min_{1\le l\le j}(\tau_l - \tau_{l+1})$.

(A.6a) The grid points $\{t_{im} : m = 1,\ldots,M\}$ are equidistant, and $n/M = O(1)$.

(A.6b) The grid points $\{t_{im} : m = 1,\ldots,M_i\}$ are independently and identically distributed with uniform density, and $\min_i M_i \asymp \max_i M_i$.

Condition (A.1) generally holds for smooth functions that are defined on finite domains. Condition (A.2) are commonly used moment conditions for $X(s,t)$. Conditions (A.3) and (A.4) are Lipschitz conditions for the joint covariance C and the marginal covariance $G_{\mathcal{S}}$ and quantify the smoothness of these covariance surfaces. Condition (A.5) requires non-zero eigengaps for the first P leading components and is widely adopted in the literature (Hall *et al.* 2006; Li and Hsing 2010). Conditions (A.6a) and (A.6b) correspond to two alternative scenarios for the design at which the underlying random process is sampled over t. Here (A.6a) reflects the case of a *dense regular* design, where one observes functions $X(\cdot,t_m)$ at a dense and regular grid of $\{t_m : m = 1,\ldots,M\}$, with $n/M = O(1)$, while (A.6b) corresponds to the case of a random design, where one observes functions $X(\cdot,t_{im})$

at a series of random locations corresponding to the time points $\{t_{im} : \ m = 1, \ldots, M_i\}$, where the number of available measurements M_i may vary across subjects.

Theorem 3. *If conditions (A.1)-(A.5),(A.6a) or (A.1)-(A.5), (A.6b) hold,* $\max(s_l - s_{l-1}) = O(n^{-1})$*, and* $\hat{\mu}(s,t)$ *obtained in Step 1 above satisfies* $\sup_{s,t} |\hat{\mu}(s,t) - \mu(s,t)| = O_p((\log n/n)^{1/2})$*, one has the following results for* $1 \leq j \leq P$*:*

$$\begin{aligned} \|\hat{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)\|_{\mathcal{S}} &= O_p((\log n/n)^{1/2}) && (15) \\ |\hat{\tau}_j - \tau_j| &= O_p((\log n/n)^{1/2}) && (16) \\ \|\hat{\psi}_j(s) - \psi_j(s)\|_{\mathcal{S}} &= O_p((\log n/n)^{1/2}) && (17) \\ \frac{1}{n}\sum_{i=1}^{n} \sup_{1\leq m\leq M_i} |\hat{\xi}_{i,j}(t_{im}) - \xi_{i,j}(t_{im})| &= O_p((\log n/n)^{1/2}). && (18) \end{aligned}$$

The *empirical estimator* and the *smoothing estimator* that are discussed in Step 1 both satisfy $\sup_{s,t} |\hat{\mu}(s,t) - \mu(s,t)| = O_p((\log n/n)^{1/2})$ under appropriate conditions and appropriate choice of the bandwidth in the *smoothing estimator*. We refer to Chen and Müller (2012), Theorems 1 and 2 for detailed conditions and proofs. The following result establishes the uniqueness of the product representation with marginal eigenfunctions ψ_j and ϕ_j derived from (5) and (14) under the common principal component assumption (13). An important implication of Theorem 4 is that the product FPCA based on marginal eigenfunctions is optimal if the eigenfunctions of kernel $C(s,t;u,v)$ indeed can be written as products in their arguments.

Theorem 4. *If there exist orthogonal bases* $\{g_j(s), j \geq 1\}$ *and* $\{f_k(t), k \geq 1\}$*, under which the common principal component assumption (13) is satisfied, we have* $g_j(s) \equiv \psi_j(s)$ *and* $f_k(t) \equiv \phi_k(t)$*, with*

$$\begin{aligned} G_{\mathcal{S}}(s,u) &= \textstyle\sum_{j=1}^{\infty} \tau_j \psi_j(s)\psi_j(u), \ \textit{for all } s,u \in \mathcal{S}, && (19) \\ G_{\mathcal{T}}(t,v) &= \textstyle\sum_{k=1}^{\infty} \vartheta_k \phi_k(t)\phi_k(v), \ \textit{for all } t,v \in \mathcal{T}, && (20) \end{aligned}$$

where

$$\begin{aligned} \tau_j &= \textstyle\sum_{k=1}^{\infty}\mathrm{var}(\chi_{jk}),\ \vartheta_k = \sum_{j=1}^{\infty}\mathrm{var}(\chi_{jk}), \\ \chi_{jk} &= \int_{\mathcal{S}}\int_{\mathcal{T}}(X(s,t)-\mu(s,t))\psi_j(s)\phi_k(t)\,dt\,ds, \\ E(\chi_{jk}) &= 0,\ \mathrm{cov}(\chi_{jk},\chi_{jl}) = \mathrm{var}(\chi_{jk})\delta_{kl},\ \mathrm{cov}(\chi_{jk},\chi_{hk}) = \mathrm{var}(\chi_{jk})\delta_{jh}. \end{aligned} \tag{21}$$

Theorem 5. *For $P \geq 1$ and $K \geq 1$, consider the following loss minimization*

$$\min_{f_k,g_j} E\left(\int_{\mathcal{S},\mathcal{T}}\{X^c(s,t) - \sum_{j=1}^{P}\sum_{k=1}^{K}\langle X^c, f_k g_j\rangle f_k(t)g_j(s)\}^2 dsdt\right),$$

with minimizing value Q^*, where f_k, $k \geq 1$ are orthogonal, and g_j, $j \geq 1$ are orthogonal. The marginal eigenfunctions $\psi_j(s)$ of $G_{\mathcal{S}}(s,u)$ and $\phi_k(t)$ of $G_{\mathcal{T}}(t,v)$ achieve good approximation in the sense that

$$E\left(\int_{\mathcal{S},\mathcal{T}}\{X^c(s,t) - \sum_{j=1}^{P}\sum_{k=1}^{K}\langle X^c, \phi_k\psi_j\rangle\phi_k(t)\psi_j(s)\}^2 dsdt\right) < Q^* + aE\|X^c\|^2,$$

where $a = \min(a_{\mathcal{T}}, a_{\mathcal{S}})$, with $(1-a_{\mathcal{T}})$ denoting the fraction of variance explained by K terms for $G_{\mathcal{T}}(t,v)$ and analogously for $a_{\mathcal{S}}$.

Similarly to the situation in Theorem 2, the error term $aE\|X^c\|^2$ depends on the loss involved in truncating just the (marginal) principal component decompositions, which also imposes a lower bound on Q^*.

5. FUNCTIONAL DATA ANALYSIS OF FERTILITY

Human fertility naturally plays a central role in demography (Preston *et al.* 2001) and its analysis recently has garnered much interest due to declining birth rates in many developed countries and associated sub-replacement fertilities (Takahashi 2004; Ezeh *et al.* 2012). The Human Fertility Database (HFD 2013) contains detailed period and cohort fertility annual data for 22 countries (plus five subdivisions: two for Germany and three for the U.K.). We are interested in *Age-Specific Fertility Rates* (ASFR), considered as functions of women's age in years (s) and repeatedly measured for each calendar year t for various countries. These rates (see (1)) constitute the functional data $X(s,t) = \mathrm{ASFR}(s,t)$.

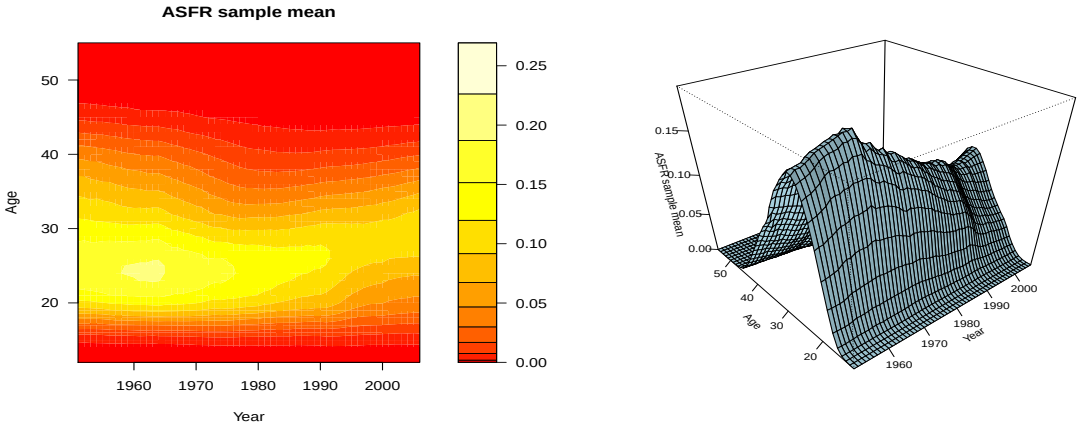

Figure 2: Sample means of the 17 fertility rate functions by calendar year.

A detailed description how ASFR is calculated from raw demographic data can be found in the HFD Methods Protocol (Jasilioniene *et al.* 2012). The specific definition of ASFR we are using corresponds to period fertility rates by calendar year and age (Lexis squares, age in completed years). In HFD (2013), $\text{ASFR}(s,t)$ is included for mothers of ages $s = 12-55$ years, thus the domain $\mathcal{S}$ is an interval of length $L = 44$ years. The interval of calendar years with available ASFR varies by country. Aiming at a compromise between the length M of the studied period $\mathcal{T}$ and the number n of countries that can be included, we choose $\mathcal{T}$ as the interval from 1951 to 2006. There are $n = 17$ countries (or territories) with available ASFR data during this time interval (see Table 4 and Figure 5 in Online Supplement B for the list of $n = 17$ included countries and heat maps depicting individual functions ASFR_i).

The sample mean $\overline{\text{ASFR}}(s,t)$ of the ASFR functions for 17 countries displayed in Figure 2 shows that fertility rates are, on average, highest for women aged between 20 and 30 and are decreasing with increasing calendar year; this overall decline is interspersed with two periods of increasing fertility before 1965, corresponding to the baby-boom, and after 1995 with a narrow increase for ages between 30 and 40 years; is narrowing in terms of the age range with high fertility; and displays an increase in regard to the ages of women where maximum fertility occurs. We applied marginal FPCA, product FPCA and two-dimensional FPCA to quantify the variability across individual countries and summarize the main results here. Additional details can be found in Online Supplement C.

The fertility data include one fertility curve over age per calendar year and per country

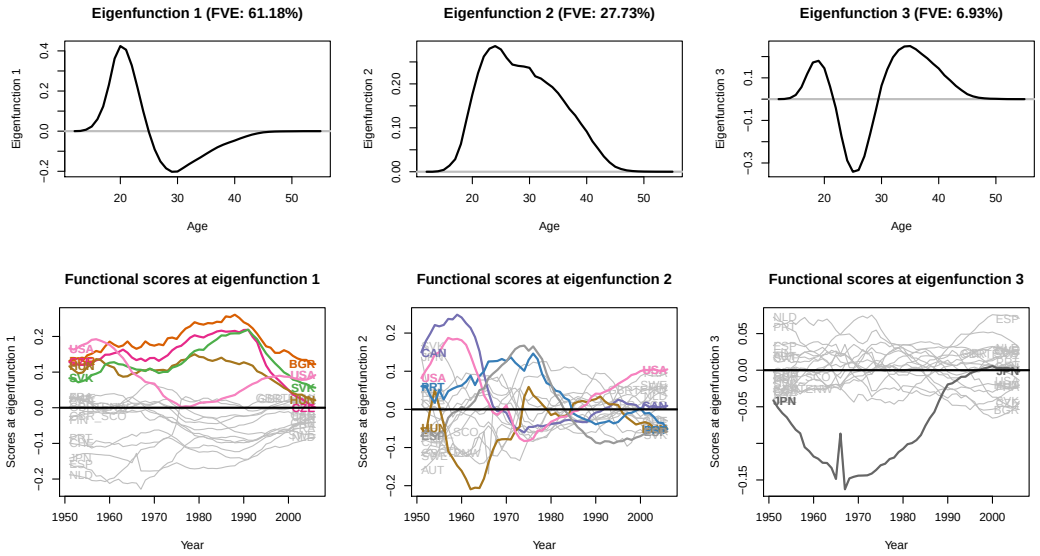

Figure 3: Results of the marginal FPCA for the fertility data. *First row:* Estimated eigenfunctions $\hat{\psi}_j(s)$, $j = 1, 2, 3$, where s is age. *Second row:* Score functions $\hat{\xi}_{i,j}(t)$, where t is calendar year. Colored lines are used for countries mentioned in the text.

and are observed on a regular grid spaced in years across both coordinates age s and calendar year t, which means that the empirical estimators described in Section 2 can be applied to these data. Figure 6 (Online Supplement B) displays the $nM = 952$ centered functional data $\text{ASFR}_i^c(s_l, t_m) = \text{ASFR}_i(s_l, t_m) - \overline{\text{ASFR}}(s_l, t_m)$, for $l = 1, \ldots, L = 44$, $m = 1, \ldots, M = 56$ and $i = 1, \ldots, n = 17$, demonstrating that there is substantial variation across countries and calendar years. The results of the proposed marginal FPCA are summarized in Figures 3 and 4 for the first three eigenfunctions, $\hat{\psi}_j(s)$, $j = 1, 2, 3$, resulting in a FVE of 95.8%. From Figure 3, the first eigenfunction $\hat{\psi}_1(s)$ can be interpreted as a contrast between fertility before and after the age of 25 years, representing the direction from mature fertility (negative scores) to young fertility (positive scores).

The second eigenfunction $\hat{\psi}_2(s)$ takes positive values for all ages s, with a maximum at age $s = 24$. The shape of $\hat{\psi}_2(s)$ is similar to that of the mean function $\overline{\text{ASFR}}(s, t)$ for a fixed year t (see the right panel of Figure 2). Therefore $\hat{\psi}_2(s)$ can be interpreted as a *size* component: Country-years with positive score in the direction of this eigenfunction have higher fertility ratios than the mean function for all ages. The third eigenfunction $\hat{\psi}_3(s)$ represents a direction from more concentrated fertility around the age of 25 years to a more

dispersed age distribution of fertility.

Examining the score functions $\hat{\xi}_{i,j}(t)$, $t \in \mathcal{T}$, which are country-specific functions of calendar year, one finds from Figure 3 for $\hat{\xi}_{i,1}(t)$ that there are countries, such as U.S. (light pink), Bulgaria (red) or Slovakia (green) for which $\xi_1(t)$ is positive for all calendar years t, which implies that these countries always have higher fertility rates for young women and vice versa for mature women, relative to the mean function. Countries from Eastern Europe such as Bulgaria, Czech Republic (pink) Hungary (brown) and Slovakia have high scores until the end of the 1980s when there is a sudden decline, implying that the relationship of fertility between younger and more mature women has reversed for these countries. Also notable is a declining trend in the dispersion of the score functions since 1990, implying that the fertility patterns of the 17 countries are converging.

The score functions $\hat{\xi}_{i,2}(t)$ corresponding to the *size* component indicate that Canada (purple) and the USA had a particularly strong baby boom in the 1960s, while Portugal (blue) and Spain (medium gray) had later baby booms during the 1970s. In contrast, Hungary had a period of relatively low fertility during the 1960s. Again, the dispersion of these size score functions declines towards 2006. The patterns of the score functions $\hat{\xi}_{i,3}(t_h)$ indicate that Japan (dark grey) has by far the largest degree of concentrated fertility at ages from 22 to 29 years, from 1960-1980, but lost this exceptional status in the 1990s and beyond. There is also a local anomaly for Japan in 1966. Takahashi (2004) reports that in 1966 the total fertility in Japan declined to the lowest value ever recorded, because 1966 was the year of the *Hinoe-Uma* (Fire Horse, a calendar event that occurs every 60 years), associated with the superstitious belief of bad luck for girls born in such years.

Trends over calendar time for particular countries can be visualized by *track plots*, which depict the changing vectors of score functions $(\xi_{i,1}(t), \ldots, \xi_{i,K}(t))$, parametrized in $t \in \mathcal{T}$, as one-dimensional curves in $\mathcal{R}^K$. Track plots are most useful for pairs of score functions and are shown in the form of planar curves for the pairs $(\xi_{i,1}(t), \xi_{i,2}(t))$ and $(\xi_{i,1}(t), \xi_{i,3}(t))$, $t \in \mathcal{T}$, in Figure 4 for selected countries and in Figure 7 (Online Supplement B) for all

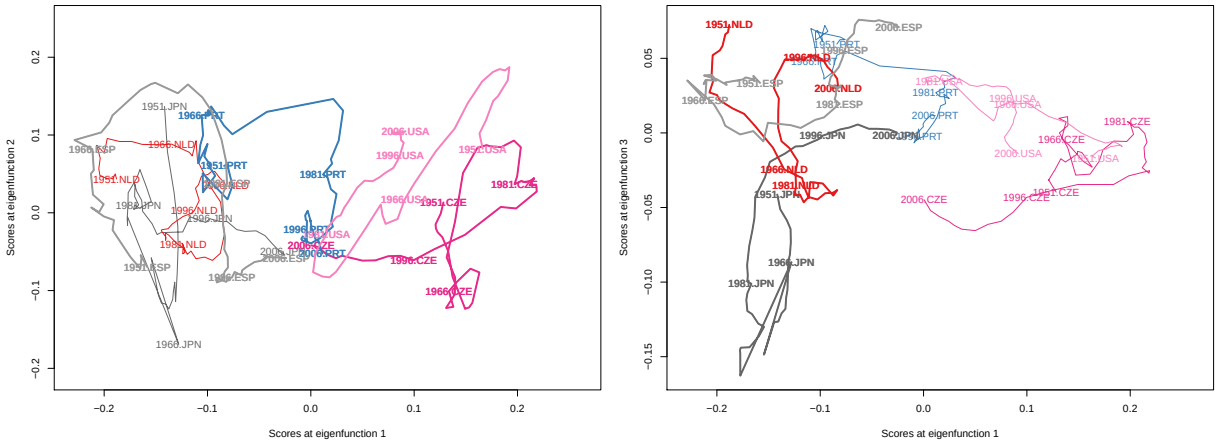

Figure 4: Track-plots $\{(\hat{\xi}_{i,1}(t), \hat{\xi}_{i,2}(t)) : t = 1951, \ldots, 2006\}$ (left panel) and $\{(\hat{\xi}_{i,1}(t), \hat{\xi}_{i,3}(t)) : t = 1951, \ldots, 2006\}$ (right panel), indexed by calendar time t, where $\xi_{i,j}(t)$ is the j-th score function for country i (for selected countries) as in (4).

countries. The left panel with the track plot illustrating the evolution in calendar time of first and second FPCs shows predominantly vertical movements: From 1951 to 2006 for most countries there are more changes in total fertility than changes in the distribution of fertility over the different ages of mothers. Exceptions to this are Portugal (blue), Spain (medium gray), Czech Republic (pink) and the U.S. (light pink), with considerable variation over the years in the first FPC score. There was more variation in fertility patterns between the countries included in this analysis in 1951 than in 2006, indicating a "globalization" of fertility patterns. In the track plot corresponding to the first and third eigenfunctions in the right panel of Figure 4, the anomalous behavior of Japan (dark gray) stands out. The third step of the marginal FPCA described in Section 2 consists of performing a separate FPCA for the estimated score functions $\hat{\xi}_{i,j}(t)$, $i = 1, \ldots, n$, for $j = 1, 2, 3$, with estimated eigenfunctions $\hat{\phi}_{jk}$ shown in Figure 8 (Online Supplement B). The interpretation of these eigenfunctions is relative to the shape of the $\psi_j(s)$.

The results in Table 1 for estimated representations (11) justify to include only the six terms with the highest FVE in the final model, leading to a cumulative FVE of 87.49%, where the FVE for each term (j, k) is estimated by (12). The corresponding 6 product functions $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$ are shown in Figure 9 (Appendix B). Regarding the comparative performance of standard two-dimensional FPCA, product FPCA (with detailed results in

Table 1: Fraction of Variance Explained (FVE) of ASFR(s,t) for the leading terms in the proposed marginal FPCA, product FPCA and two-dimensional FPCA. Number of terms in each case is selected to achieve fraction of variance explained (FVE) of more than 85%.

marginal FPCA	FVE in %	product FPCA	FVE in %	2d FPCA	FVE in %
Six terms	87.49	Seven terms	87.38	Four terms	89.73
$\hat{\phi}_{11}(t)\hat{\psi}_1(s)$	54.33	$\hat{\phi}_1(t)\hat{\psi}_1(s)$	53.69	$\hat{\gamma}_1(s,t)$	58.93
$\hat{\phi}_{21}(t)\hat{\psi}_2(s)$	13.04	$\hat{\phi}_2(t)\hat{\psi}_2(s)$	8.10	$\hat{\gamma}_2(s,t)$	13.71
$\hat{\phi}_{22}(t)\hat{\psi}_2(s)$	6.88	$\hat{\phi}_1(t)\hat{\psi}_2(s)$	8.08	$\hat{\gamma}_3(s,t)$	11.04
$\hat{\phi}_{12}(t)\hat{\psi}_1(s)$	4.62	$\hat{\phi}_3(t)\hat{\psi}_2(s)$	5.51	$\hat{\gamma}_4(s,t)$	6.05
$\hat{\phi}_{23}(t)\hat{\psi}_2(s)$	4.40	$\hat{\phi}_2(t)\hat{\psi}_1(s)$	4.47		
$\hat{\phi}_{31}(t)\hat{\psi}_3(s)$	4.22	$\hat{\phi}_4(t)\hat{\psi}_2(s)$	3.85		
		$\hat{\phi}_1(t)\hat{\psi}_3(s)$	3.68		

Online Supplement C) and marginal FPCA, we find: (1) As expected, standard FPCA based on the two-dimensional Karhunen-Loève expansion requires fewer components to explain a given amount of variance, as 4 eigenfunctions lead to a FVE of 89.73% (see Table 1), while marginal FPCA representation achieves a FVE of 87.49% with 6 terms, and product FPCA needs 7 terms to explain 87.38%. (2) Product FPCA and Marginal FPCA represent the functional data as a sum of terms that are products of two functions, each depending on only one argument. This provides for much better interpretability and makes it possible to discover patterns in functional data that are not found when using standard FPCA. For instance, the second eigenfunction ψ_2 in the first step of the marginal FPCA could be characterized as a *fertility size* component, with a country-specific time-varying multiplier $\xi_2(t)$. Standard FPCA does not pinpoint this feature, which is an essential characteristic of demographic changes in fertility. (3) Marginal FPCA makes it much easier than standard FPCA to analyze the time dynamics of the fertility process.

Specifically, the plots in the second row of Figure 3 or the track plots in Figure 4 are informative about the fertility evolution over calendar years: (a) The relative balance between young and mature fertility at each country changes over the years. The graphical

representation of functional score functions $\hat{\xi}_{i,1}(t)$ allows to characterize and quantify this phenomenon. (b) The track plot in the left panel of Figure 4 indicates that in general it is much more common that fertility rates rise or decline across all ages compared to transfers of fertility between different age groups. (c) The fertility patterns of the various countries are much more similar in 2006 than in 1951.

All three approaches to FPCA for function-valued stochastic processes, namely standard FPCA (3), marginal FPCA (4) and the product FPCA (6), can be used to produce country scores which can be plotted against each other. They turn out to be similar for these approaches; as an example the standard FPCA scores are shown in Figure 12 (Appendix C). We conclude that standard FPCA, marginal FPCA and product FPCA complement each other. Our recommendation is to perform all whenever feasible, in order to gain as much insight about complex functional data as possible.

6. SIMULATIONS

We conducted two simulation studies, one to investigate the estimation procedure for marginal FPCA, and a second study to evaluate the performance of product FPCA. Both were conducted in a scenario that mimicks the fertility data. For simulation 1, we generated data following a truncated version of (4), where we used the estimated mean function $\overline{ASFR}(s,t)$ from the country fertility data (Section 5) as mean function and the estimated product functions $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$, $1 \leq j,k \leq 4$, as base functions in (4). Random scores χ_{jk} were generated as independent normal random variables with variances λ_{jk}, corresponding to the estimates derived from the fertility data, $\lambda_{jk} = \frac{1}{n}\sum_{i=1}^{n} \hat{\chi}_{i,jk}^2$. We also added i.i.d. noise to the actual observations $Y_i(s_l,t_h) = X(s_l,t_h) + \epsilon_{i,lh}$, $l = 1,\ldots,44$, $h = 1,\ldots,56$, where $\epsilon_{i,lh} \sim N(0,\sigma^2)$ with $\sigma = 0.005$ to mimic the noise level of the fertility data.

Estimated and true functions $\psi_j(s)$ and $\phi_{jk}(t)$ obtained for one sample run with $n = 50$ are shown in Figure 10 (Online Supplement B), demonstrating very good recovery of the true basis functions. To quantify the quality of the estimates of $\mu(s,t)$, we use the relative

Table 2: Results for simulation 1, reporting median relative errors (RE), as defined in (22) (with median absolute deviation in parentheses), for various components of the model and varying sample sizes n.

RE	FVE in %	$n = 50$	$n = 100$	$n = 200$
μ		0.0012 (0.0008)	0.0006 (0.0004)	0.0003 (0.0002)
X^c		0.1523 (0.0228)	0.1483 (0.0168)	0.1435 (0.0091)
$\phi_{11}(t)\psi_1(s)$	53.6967	0.0092 (0.0071)	0.0045 (0.0040)	0.002 (0.0016)
$\phi_{21}(t)\psi_2(s)$	12.9333	0.0584 (0.0538)	0.0280 (0.0243)	0.0133 (0.0110)
$\phi_{22}(t)\psi_2(s)$	6.7450	0.1306 (0.1267)	0.0660 (0.0619)	0.0311 (0.0287)
$\phi_{12}(t)\psi_1(s)$	4.5367	0.0222 (0.0178)	0.0129 (0.0091)	0.005 (0.0037)
$\phi_{23}(t)\psi_2(s)$	4.1917	0.0999 (0.0904)	0.0469 (0.0417)	0.0296 (0.0238)
$\phi_{31}(t)\psi_3(s)$	4.0400	0.0283 (0.0238)	0.0127 (0.0100)	0.0077 (0.0062)

squared error

$$RE = \frac{||\mu(s,t) - \hat{\mu}(s,t)||^2}{||\mu(s,t)||^2}, \tag{22}$$

where $||\mu(s,t)||^2 = \int\int \mu(s,t)^2 dsdt$, analogously for $\hat{X}_i^c(s,t)$ and $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$. The relative squared errors over 200 simulation runs, reported in Table 2, were found to be quite small for μ, X_i^c and for the six product functions $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$ with largest FVEs, which are the same six functions as in Figure 9. The errors decline with increasing sample size n, as expected. The FVEs for each term (j,k) are also in Table 2, averaged over simulation runs and over the different sample sizes, as they were similar across varying sample sizes.

For simulation 2, data were generated according to a truncated product FPC model

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{4}\sum_{k=1}^{4} \chi_{jk}\phi_k(t)\psi_j(s),$$

where $\mu(s,t)$ and $\phi_k(t)\psi_j(s)$ for $1 \leq j,k \leq 4$ are substituted by the estimates obtained from the fertility data. As in simulation 1, the random scores χ_{jk} were generated as independent normal random variables with variances estimated from the data. Estimated and true functions $\psi_j(s)$ and $\phi_k(t)$ obtained for one sample run with $n = 50$ are shown in Figure 11 (Online Supplement B). The relative squared errors over 200 simulation runs, for μ, X_i^c

Table 3: Results for simulation 2, reporting median relative errors (RE), as defined in (22) (with median absolute deviations in parentheses) for various components of the model and varying sample sizes n.

RE	FVE in %	$n = 50$	$n = 100$	$n = 200$
μ		0.0011 (0.0006)	0.0006 (0.0003)	0.0003 (0.0002)
X^c		0.1464 (0.0216)	0.1390 (0.0138)	0.1324 (0.0080)
$\phi_1(t)\psi_1(s)$	53.5400	0.0099 (0.0075)	0.0053 (0.0048)	0.0022 (0.0017)
$\phi_2(t)\psi_2(s)$	8.1500	0.0559 (0.0525)	0.0342 (0.0275)	0.0174 (0.0180)
$\phi_1(t)\psi_2(s)$	8.0817	0.0109 (0.0078)	0.0064 (0.0051)	0.0026 (0.0018)
$\phi_3(t)\psi_2(s)$	5.4300	0.0776 (0.0635)	0.0389 (0.0331)	0.0208 (0.0204)
$\phi_2(t)\psi_1(s)$	4.2783	0.0543 (0.0502)	0.0328 (0.0271)	0.0173 (0.0180)
$\phi_4(t)\psi_2(s)$	3.7817	0.0368 (0.0293)	0.0167 (0.0131)	0.0089 (0.0072)
$\phi_1(t)\psi_3(s)$	3.5917	0.0108 (0.0075)	0.0056 (0.0039)	0.0028 (0.0019)

and for the seven product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$ with largest FVEs (among 16 total product functions), which are the same seven functions as in Figure 18 (Online Supplement C), are reported in Table 3. Both figure and numbers demonstrate good performance of the method.

7. DISCUSSION

The proposed marginal FPCA and product FPCA provide a simple and straightforward representation of function-valued stochastic processes. This holds especially in comparison with a previously proposed two-step expansion for repeatedly observed functional data (Chen and Müller 2012), in which processes X are represented as

$$X(s,t) = \mu(s,t) + \sum_{j=1}^{\infty} \nu_j(t)\rho_j(s|t) = \mu(s,t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \theta_{jk}\omega_{jk}(t)\rho_j(s|t), \tag{23}$$

where $\rho_j(\cdot|t)$ is the j-th eigenfunction of the operator in $L^2(\mathcal{S})$ with kernel $G_{\mathcal{S}}(s,u|t) = C((s,t),(u,t))$, $\nu_j(t) = \langle X(\cdot,t), \rho_j(\cdot|t)\rangle_{\mathcal{S}}$ and $\sum_{k=1}^{\infty}\theta_{jk}\omega_{jk}(t)$ is the Karhunen-Loève expansion of $\nu_j(t)$ as a random function in $L^2(\mathcal{T})$. This method can be characterized as a

conditional FPCA approach (we note that in Chen and Müller (2012) the notation of s and t is reversed as compared to the present paper). Similarly to the proposed marginal approach this conditional method provides for asymmetric handling of arguments s and t of X and is a two-step procedure which is composed of iterated one-dimensional FPCAs.

Key differences between the marginal FPCA and the conditional FPCA are as follows: (1) The first step of the conditional FPCA approach (23) requires to perform a separate FPCA for each $t \in \mathcal{T}$, while in the marginal approach (4) only one FPCA is required, with lower computational cost, and, most importantly, using all the data rather than the data in a window around t. (2) In (23), the eigenfunctions $\rho_j(s|t)$ depend on both arguments, making it difficult to separate and interpret the effects of s and t in conditional FPCA, in contrast to marginal FPCA, where the eigenfunctions in (4) only depend on s. (3) For sparse designs, conditional FPCA requires a smoothing estimator of the function $G_{\mathcal{S}}(s, u|t)$ that depends on $2d_1 + d_2$ univariate arguments. This improves upon the standard two-argument FPCA (3), where the corresponding covariance functions depend on $2d_1 + 2d_2$ arguments. The improvement is however even greater for marginal FPCA, where the covariance function depends on only $2d_1$ or $2d_2$ arguments, leading to faster convergence.

The proposed marginal FPCA improves upon standard FPCA by providing interpretable components and making it possible to treat the index of the stochastic process asymmetrically in the arguments of the random functions that constitute the values of the process. While we have discussed in detail the case of time-indexed function-valued processes, and our example also conforms with this simplest setting, extensions to spatially indexed function-valued processes or processes which are indexed on a rectangular subdomain of $\mathcal{R}^p$ are straightforward. Marginal FPCA also is supported by theoretical optimality properties as per Theorem 1 and Theorem 2.

A promising simplified version of the marginal FPCA is product FPCA, motivated by a common principal component assumption, see Theorem 4. Additional motivation is its near optimality even without the common principal component assumption, as per Theorem 5.

In our fertility data example, the loss of flexibility is quite limited and may be outweighed by the simplicity and interpretability of this model. In general, the explanatory power of the product FPCA model depends on the structure of the double-indexed array $\eta_{jk} = \text{var}(\chi_{jk})$. When one of the marginal kernels does not have fast decaying eigenvalues, relatively large values of η_{jk} might show up in large j or large k and in such situations the product FPCA might have limited explanatory power, and it would be better to apply marginal FPCA or two-dimensional FPCA. The eigenvalues of the marginal kernels can be directly estimated and can be used to diagnose this situation in data applications.

In this paper we mainly focus on the case where the argument of the functional values s is densely and regularly sampled. In practical applications with designs that are sparse in s, one may obtain $\hat{G}_{\mathcal{S}}$ by pooling the data $\{\hat{X}_i^c(\cdot, t_{im}),\ i = 1, \ldots, n,\ m = 1, \ldots, M_i\}$, and utilizing a two-dimensional smoothing estimator of the covariance (Yao *et al.* 2005). The FPCs can be obtained through conditional expectation (PACE) under Gaussian assumptions; software is available at `http://www.stat.ucdavis.edu/PACE/`. For this case, one can only show that $\hat{\xi}_{i,j}(t) \rightarrow_{a.s.} E(\xi_{i,j}(t)|\text{Data})$ under Gaussian assumptions.

REFERENCES

Aston, J. A., Pigoli, D. and Tavakoli, S. (2015) Tests for separability in nonparametric covariance operators of random surfaces. *arXiv preprint arXiv:1505.02023*.

Bosq, D. (2000) *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer.

Chen, K. and Müller, H.-G. (2012) Modeling repeated functional observations. *Journal of the American Statistical Association*, **107**, 1599–1609.

Crainiceanu, C., Staicu, A. and Di, C. (2009) Generalized multilevel functional regression. *Journal of the American Statistical Association*, **104**, 1550–1561.

Cuevas, A. (2013) A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, **147**, 1–23.

Currie, I. D., Durban, M. and Eilers, P. H. (2004) Smoothing and forecasting mortality rates. *Statistical modelling*, **4**, 279–298.

— (2006) Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 259–280.

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010) Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21**, 224–239.

Eilers, P. H., Currie, I. D. and Durbán, M. (2006) Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis*, **50**, 61–76.

Eilers, P. H. and Marx, B. D. (2003) Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems*, **66**, 159–174.

Ezeh, A. C., Bongaarts, J. and Mberu, B. (2012) Global population trends and policy options. *The Lancet*, **380**, 142–148.

Greven, S., Crainiceanu, C., Caffo, B., Reich, D. *et al.* (2010) Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, **4**, 1022–1054.

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2012) Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *The Annals of Applied Statistics*, **6**, 669–696.

— (2013) Nonparametric estimation in small data sets of spatially indexed curves with application to temporal trend determination. *Computational Statistics and Data Analysis*, **59**, 82–94.

Hall, P., Müller, H.-G. and Wang, J.-L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, **34**, 1493–1517.

HFD (2013) Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). (Version March 18, 2013). URL http://www.humanfertility.org.

Horváth, L. and Kokoszka, P. (2012) *Inference for Functional Data with Applications*. New

York: Springer.

Huang, J. Z., Shen, H. and Buja, A. (2009) The analysis of two-way functional data using two-way regularized singular value decompositions. *Journal of the American Statistical Association*, **104**, 1609–1620.

Hyndman, R. J. and Shang, H. L. (2009) Forecasting functional time series. *Journal of the Korean Statistical Society*, **38**, 199–211.

Hyndman, R. J. and Ullah, S. (2007) Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, **51**, 4942–4956.

Jasilioniene, A., Jdanov, D. A., Sobotka, T., Andreev, E. M., Zeman, K. and Shkolnikov, V. M. (2012) *Methods Protocol for the Human Fertility Database*. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria).

Kneip, A. and Utikal, K. (2001) Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, **96**, 519–542.

Li, Y. and Hsing, T. (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics*, **38**, 3321–3351.

Morris, J. S. and Carroll, R. J. (2006) Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B*, **68**, 179–199.

Nerini, D., Monestiez, P. and Manté, C. (2010) Cokriging for spatial functional data. *Journal of Multivariate Analysis*, **101**, 409–418.

Park, S. and Staicu, A. (2015) Longitudinal functional data analysis. *STAT*, **4**, 212–226.

Preston, S. H., Heuveline, P. and Guillot, M. (2001) *Demography: Measuring and modeling population processes*. Blackwell Publishing.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. Springer Series in Statistics. New York: Springer, second edn.

Takahashi, S. (2004) Demographic investigation of the declining fertility process in japan. *The Japanese Journal of Population*, **2**, 93–116.

WPP (2012) World population prospects: The 2012 revision, dvd edition.

Yao, F. and Lee, T. C. M. (2006) Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society: Series B*, **68**, 3–25.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590.

Yuan, Y., Gilmore, J. H., Geng, X., Martin, S., Chen, K., Wang, J.-l. and Zhu, H. (2014) Fmem: Functional mixed effects modeling for the analysis of longitudinal white matter tract data. *NeuroImage*, **84**, 753–764.

Online Supplement

Supplement A: Auxiliary Results and Proofs

Lemma 1. *The function $G_{\mathcal{S}}$ defined in (5) is a continuous, symmetric and positive definite function. Moreover it is a valid covariance function.*

Proof. Symmetry is obvious. Continuity follows from the continuity of the covariance function C. We have that $G_{\mathcal{S}}(s,u) = \int_{\mathcal{T}} \text{cov}(X(s,t), X(u,t))dt$. Observing that for each fixed t, $\text{cov}(X(s,t), X(u,t))$ is positive definite, it holds that for any function f in $L^2(\mathcal{S})$ we have that $\int_{\mathcal{S}\times\mathcal{S}} \text{cov}(X(s,t), X(u,t))f(u)f(s)duds \geq 0$ for all $t \in \mathcal{T}$. Therefore, by Fubini,

$$\int_{\mathcal{S}\times\mathcal{S}} G_{\mathcal{S}}(u,s)f(u)f(s)duds = \int_{\mathcal{T}}\int_{\mathcal{S}\times\mathcal{S}} \text{cov}(X(s,t), X(u,t))f(u)f(s)dudsdt \geq 0.$$

To see that $G_{\mathcal{S}}$ is a valid covariance function, remember that this is equivalent to say that $\int_{\mathcal{S}} G_{\mathcal{S}}(s,s)ds < \infty$ (see, for instance, Horváth and Kokoszka 2012, page 24). Observe that

$$\int_{\mathcal{S}} G_{\mathcal{S}}(s,s)ds = \int_{\mathcal{S}}\int_{\mathcal{T}} C((s,t),(s,t))dtds$$

and the last integral is finite because C is a valid covariance function. □

Lemma 2. *Let $\xi_j(t) = \langle X(\cdot,t) - \mu(\cdot,t), \psi_j\rangle_{\mathcal{S}}$, $j \geq 1$, be the random functional coefficients in the marginal Karhunen-Loève representation (7). Then $E(\xi_j(t)) = 0$ for almost all $t \in \mathcal{T}$, and $E(\langle \xi_j, \xi_k\rangle_{\mathcal{T}}) = \tau_j\delta_{jk}$, where $\delta_{jk} = 1$ if $j = k$ and $= 0$ otherwise.*

Proof. First, for almost all $t \in \mathcal{T}$, $X(\cdot,t)$ is a random element of $L^2(\mathcal{S})$ because X is in $L^2(\mathcal{S}\times\mathcal{T})$. Then there exists a unique (in the L^2 sense) function $\tilde{\mu}(\cdot,t)$ in $L^2(\mathcal{S})$ such that $E(\langle X(\cdot,t), y\rangle_{\mathcal{S}}) = \langle \tilde{\mu}(\cdot,t), y\rangle_{\mathcal{S}}$ for all $y \in L^2(\mathcal{S})$ and it follows that $\tilde{\mu}(s,t) = E(X(s,t)) = \mu(s,t)$ for almost all $s \in \mathcal{S}$ (see, for instance, Horváth and Kokoszka (2012), Section 2.3), so that $\tilde{\mu}(\cdot,t) = \mu(\cdot,t)$ in the L^2 sense. Then taking $y = \psi_j$ we have that

$$E(\xi_j(t)) = E\left(\langle X(\cdot,t), \psi_j\rangle_{\mathcal{S}}\right) - \langle \mu(\cdot,t), \psi_j\rangle_{\mathcal{S}} = 0.$$

Furthermore,

$$
\begin{aligned}
&E(\langle \xi_j, \xi_k \rangle_{\mathcal{T}}) = E\left(\int_{\mathcal{T}} \xi_j(t)\xi_k(t)dt \right) \\
=&E\left(\int_{\mathcal{T}} \left(\int_{\mathcal{S}} X^c(s,t)\psi_j(s)ds \right) \left(\int_{\mathcal{S}} X^c(u,t)\psi_k(u)du \right) dt \right) \\
=&\int_{\mathcal{S}}\int_{\mathcal{S}}\int_{\mathcal{T}} E\left(X^c(s,t)X^c(u,t)\right) dt\ \psi_j(s)\psi_k(u)\ ds\ du \\
=&\int_{\mathcal{S}} \left(\int_{\mathcal{S}} G_{\mathcal{S}}(s,u)\psi_j(s)\ ds \right) \psi_k(u)\ du \\
=&\langle \Psi(\psi_j), \psi_k \rangle_{\mathcal{S}} = \tau_j \langle \psi_j, \psi_k \rangle_{\mathcal{S}} = \tau_j \delta_{jk},
\end{aligned}
$$

where we have used that τ_j, ψ_j, $j \geq 1$, are, respectively, the eigenvalues and eigenfunctions of the operator Ψ defined as $\Psi(f)(s) = \int G_{\mathcal{S}}(s,u)\, f(u)\, du$. □

Proof of Theorem 1:

Observe that

$$
\begin{aligned}
&E\left(\int_{\mathcal{T}} \|X^c(\cdot,t) - \sum_{j=1}^{P} \langle X^c(\cdot,t), g_j \rangle_{\mathcal{S}} g_j \|_{\mathcal{S}}^2 dt \right) \\
=&E\left(\int_{\mathcal{T}} \langle X^c(\cdot,t) - \sum_{j=1}^{P} \langle X^c(\cdot,t), g_j \rangle_{\mathcal{S}} g_j, X^c(\cdot,t) - \sum_{j=1}^{P} \langle X^c(\cdot,t), g_j \rangle_{\mathcal{S}} g_j \rangle_{\mathcal{S}} dt \right) \\
=&E\left(\int_{\mathcal{T}} \langle X^c(\cdot,t), X^c(\cdot,t) \rangle_{\mathcal{S}} dt \right) - \sum_{j=1}^{P} E\left(\int_{\mathcal{T}} \langle X^c(\cdot,t), g_j \rangle_{\mathcal{S}}^2 dt \right) \\
=&E\left(\int_{\mathcal{T}} \int_{\mathcal{S}} \left(X^c(s,t)\right)^2 ds dt \right) - \sum_{j=1}^{P} E\left(\int_{\mathcal{T}} \left(\int_{\mathcal{S}} X^c(s,t) g_j(s) ds \right)^2 dt \right) \\
=&E\left(\|X^c\|^2 \right) - \sum_{j=1}^{P} \int_{\mathcal{S}}\int_{\mathcal{S}}\int_{\mathcal{T}} E\left(X^c(s,t)X^c(u,t)\right) dt\ g_j(u) du\ g_j(s) ds \\
=&E\left(\|X^c\|^2 \right) - \sum_{j=1}^{P} \int_{\mathcal{S}}\int_{\mathcal{S}}\int_{\mathcal{T}} C((s,t),(u,t)) dt\ g_j(u) du\ g_j(s) ds \\
=&E\left(\|X^c\|^2 \right) - \sum_{j=1}^{P} \int_{\mathcal{S}}\int_{\mathcal{S}} G_{\mathcal{S}}(s,u) g_j(u) du\ g_j(s) ds \\
=&E\left(\|X^c\|^2 \right) - \sum_{j=1}^{P} \langle \Psi(g_j), g_j \rangle_{\mathcal{S}},
\end{aligned}
$$

where Ψ is the operator in $L^2(\mathcal{S})$ with kernel $G_{\mathcal{S}}$. Then minimizing

$$E\left(\int_{\mathcal{T}} \|X^c(\cdot,t) - \sum_{j=1}^{P} \langle X^c(\cdot,t), g_j\rangle_{\mathcal{S}} g_j\|_{\mathcal{S}}^2 dt\right)$$

is equivalent to maximizing $\sum_{j=1}^{P} \int_{\mathcal{S}} \langle \Psi(g_j), g_j\rangle_{\mathcal{S}}$. Given that Ψ is a symmetric, positive definite Hilbert-Schmidt operator (see Lemma 1), standard arguments (see, for instance, Theorem 3.2 in Horváth and Kokoszka (2012)) complete the proof.

Proof of Theorem 2:

For $f_{jk}(t)$ and $g_j(s)$ that satisfy the orthogonal conditions, we have

$$\begin{aligned}
&E\left(\int_{\mathcal{S},\mathcal{T}} \{X^c(s,t) - \sum_{j=1}^{P}\sum_{k=1}^{K_j} \langle X^c, f_{jk}g_j\rangle f_{jk}(t)g_j(s)\}^2 dsdt\right) \\
=&E\|X^c\|^2 - 2\times \sum_{j=1}^{P}\sum_{k=1}^{K_j} E\left(\int_{\mathcal{S},\mathcal{T}} X^c(s,t)\langle X^c, f_{jk}g_j\rangle f_{jk}(t)g_j(s)\right) dsdt \\
&+ E\left(\int_{\mathcal{S},\mathcal{T}} \sum_{j=1}^{P}\sum_{k=1}^{K_j}\sum_{l=1}^{P}\sum_{h=1}^{K_j} \langle X^c, f_{jk}g_j\rangle f_{jk}(t)g_j(s)\langle X^c, f_{lh}g_l\rangle f_{lh}(t)g_l(s) dsdt\right) \\
=&E\|X^c\|^2 - 2\times \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(X^c(s,t)X^c(u,v)\right) f_{jk}(t)f_{jk}(v)g_j(s)g_j(u) dsdtdudv \\
&+ \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(X^c(s,t)X^c(u,v)\right) f_{jk}(t)f_{jk}(v)g_j(s)g_j(u) dsdtdudv \\
=&E\|X^c\|^2 - \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(X^c(s,t)X^c(u,v)\right) f_{jk}(t)f_{jk}(v)g_j(s)g_j(u) dsdtdudv
\end{aligned} \tag{24}$$

Let $\tilde{f}_{jk}(t)$ and $\tilde{g}_j(s)$ denote the optimal basis to achieve the minimum reconstruction error Q^*, and define

$$(I) = \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(X^c(s,t)X^c(u,v)\right) \phi_{jk}(t)\phi_{jk}(v)\psi_j(s)\psi_j(u) dsdtdudv$$

and

$$(II) = \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(X^c(s,t)X^c(u,v)\right) \tilde{f}_{jk}(t)\tilde{f}_{jk}(v)\tilde{g}_j(s)\tilde{g}_j(u) dsdtdudv.$$

By Eq. (24), to prove the theorem, we only need to show that $(II) - (I) < aE\|X^c\|_2$.

We further define

$$(III) = \sum_{j=1}^{P} \int_{\mathcal{S}\times\mathcal{S}} \int_{\mathcal{T}} E\left(X^c(s,t)X^c(u,t)\right) dt\tilde{g}_j(s)\tilde{g}_j(u)dsdu,$$

and

$$(IV) = \sum_{j=1}^{P} \int_{\mathcal{S}\times\mathcal{S}} \int_{\mathcal{T}} E\left(X^c(s,t)X^c(u,t)\right) dt\psi_j(s)\psi_j(u)dsdu.$$

We will prove that $(II) < (III) < (IV)$ and $(IV) - (I) < aE\|X^c\|_2$, which implies that $(II) - (I) < aE\|X^c\|_2$.

By definition, ψ_j are the leading eigenfunctions of the marginal kernel $G_{\mathcal{S}}(s,u)$ so that $(III) < (IV)$.

To show $(II) < (III)$, let $\tilde{\xi}_j(t) = \langle X^c(s,t), \tilde{g}_j(s)\rangle$. Then,

$$\begin{aligned}
(II) =& \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(X^c(s,t)X^c(u,v)\right) \tilde{f}_{jk}(t)\tilde{f}_{jk}(v)\tilde{g}_j(s)\tilde{g}_j(u)dsdtdudv \\
=& \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(\tilde{\xi}_j(t)\tilde{\xi}_j(v)\right) \tilde{f}_{jk}(t)\tilde{f}_{jk}(v)dtdv \\
<& \sum_{j=1}^{P}\sum_{k=1}^{\infty} \int E\left(\tilde{\xi}_j(t)\tilde{\xi}_j(v)\right) \tilde{f}_{jk}(t)\tilde{f}_{jk}(v)dtdv \\
=& \sum_{j=1}^{P} \int E\left(\tilde{\xi}_j(t)\tilde{\xi}_j(t)\right) dt \\
=& \sum_{j=1}^{P} \int E\left(\int X^c(s,t)\tilde{g}_j(s)ds \int X^c(u,t)\tilde{g}_j(u)du\right) dt \\
=& (III)
\end{aligned} \tag{25}$$

Finally, we prove $(IV) - (I) < aE\|X^c\|_2$,

$$\begin{aligned}
(IV) - (I) =& \sum_{j=1}^{P} \int E\left(\xi_j(t)\xi_j(t)\right) dt \\
& - \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(X^c(s,t)X^c(u,v)\right) \phi_{jk}(t)\phi_{jk}(v)\psi_j(s)\psi_j(u) ds dt du dv \\
=& \sum_{j=1}^{P}\sum_{k=1}^{\infty} \int E\left(\xi_j(t)\xi_j(v)\right) \phi_{jk}(t)\phi_{jk}(v) dt dv \\
& - \sum_{j=1}^{P}\sum_{k=1}^{K_j} \int E\left(\xi_j(t)\xi_j(v)\right) \phi_{jk}(t)\phi_{jk}(v) dt dv \\
& < aE\|X^c\|_2,
\end{aligned} \tag{26}$$

where $a = \max_{1\le j\le P} a_j$, with $(1-a_j)\%$ denoting the fraction of variance explained by K_j terms in each process $\xi_j(t) = \langle X^c(\cdot,t), \psi_j\rangle$.

Proof of Theorem 3:

Recall that

$$G_{\mathcal{S}}(s,u) = \int_{\mathcal{T}} C((s,t),(u,t))dt,$$

where $C((s,t),(u,t)) = E[(X(s,t)-\mu(s,t))(X(u,t)-\mu(u,t))]$. For (s,u) on the grid points, we have

$$\hat{G}_{\mathcal{S}}(s,u) = \frac{|\mathcal{T}|}{\sum_{i=1}^{n} M_i} \sum_{i=1}^{n}\sum_{m=1}^{M_i} \hat{X}^c(s,t_{im})\hat{X}^c(u,t_{im}),$$

where $\hat{X}^c(s,t_{im}) = X(s,t_{im}) - \hat{\mu}(s,t_{im})$ and $|\mathcal{T}|$ is the Lebesgue measure of $\mathcal{T}$. We define

$$\tilde{G}_{\mathcal{S}}(s,u) = \frac{|\mathcal{T}|}{\sum_{i=1}^{n} M_i} \sum_{i=1}^{n}\sum_{m=1}^{M_i} X^c(s,t_{im})X^c(u,t_{im}).$$

Using $\sup_{s,t}|\hat{\mu}(s,t) - \mu(s,t)| = O_p((\log n/n)^{1/2})$, it is easy to show that $\|\tilde{G}_{\mathcal{S}}(s,u) - \hat{G}_{\mathcal{S}}(s,u)\|_{\mathcal{S}} = O_p((\log n/n)^{1/2})$. Next we show

$$\|\tilde{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)\|_{\mathcal{S}} = O_p((1/n)^{1/2}). \tag{27}$$

We first prove (27) under assumption (A.6a). By (A.6a), we have $M_i \equiv M$, and the grid of t is $\{t_1, \ldots, t_M\}$. Therefore,

$$\begin{aligned} &\sup_{(s,u)\in\mathcal{S}^2} |E\tilde{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)| \\ &= \sup_{(s,u)\in\mathcal{S}^2} \left| \frac{|\mathcal{T}|}{M} \sum_{m=1}^{M} C((s,t_m),(u,t_m)) - \int_{\mathcal{T}} C((s,t),(u,t))dt \right| \\ &\overset{A.3}{=} O(\frac{1}{M}) = O(1/n), \end{aligned} \tag{28}$$

and

$$\begin{aligned} &\sup_{(s,u)\in\mathcal{S}^2} \mathrm{var}(\tilde{G}_{\mathcal{S}}(s,u)) \\ &= \sup_{(s,u)\in\mathcal{S}^2} \frac{|\mathcal{T}|^2}{(nM)^2} \sum_{i=1}^{n} \mathrm{var}(\sum_{m=1}^{M} X^c(s,t_m)X^c(u,t_m)) \\ &\leq \sup_{(s,u)\in\mathcal{S}^2} \frac{|\mathcal{T}|^2}{(nM)^2} \sum_{i=1}^{n} \sum_{m,m'=1}^{M} E(X^c(s,t_m)X^c(u,t_m)X^c(s,t_{m'}))X^c(u,t_{m'})) \\ &\overset{A.1,A.2}{\leq} \frac{|\mathcal{T}|^2}{n^2M^2} \sum_{i=1}^{n} M^2B = O(1/n). \end{aligned} \tag{29}$$

Combining (28) and (29) we have

$$\sup_{(s,u)\in\mathcal{S}^2} E|\tilde{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)|^2 = O(1/n).$$

Therefore, by (A.4) and $(s_l - s_{l-1}) = O(n^{-1})$,

$$\begin{aligned} E\|\tilde{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)\|_{\mathcal{S}}^2 &= \int_{\mathcal{S}}\int_{\mathcal{S}} |\tilde{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)|^2 dsdu \\ &= \frac{|\mathcal{S}|}{L^2} \sum_{j=1}^{L}\sum_{l=1}^{L} E|\tilde{G}_{\mathcal{S}}(s_j,s_l) - G_{\mathcal{S}}(s_j,s_l)|^2 + O(1/n) \\ &= O(1/n), \end{aligned} \tag{30}$$

which implies (27). The same result can be derived using a similar argument under (A.6b),

by substituting (28) with

$$\sup_{(s,u)\in\mathcal{S}^2} |E\tilde{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)| \tag{31}$$
$$= \left| E\frac{|\mathcal{T}|}{\sum_{i=1}^n M_i} \sum_{i=1}^n \sum_{m=1}^{M_i} C((s,t_{im}),(u,t_{im})) - \int_{\mathcal{T}} C((s,t),(u,t))dt \right| = 0,$$

and substituting (29) with

$$\sup_{(s,u)\in\mathcal{S}^2} \mathrm{var}(\tilde{G}_{\mathcal{S}}(s,u)) = \frac{|\mathcal{T}|^2}{(\sum_{i=1}^n M_i)^2} \sum_{i=1}^n \mathrm{var}(\sum_{m=1}^{M_i} X^c(s,t_{im})X^c(u,t_{im})) \tag{32}$$
$$\overset{A.1,A.2}{\leq} \frac{|\mathcal{T}|^2}{(\sum_{i=1}^n M_i)^2} \sum_{i=1}^n M_i^2 B = O(1/n).$$

This completes the proof for Eq. (15).

For a fixed j, Lemma 4.3 in Bosq (2000) implies that

$$|\hat{\tau}_j - \tau_j| \leq ||\hat{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)||_{\mathcal{S}}, \quad ||\hat{\psi}_j(s) - \psi_j(s)||_{\mathcal{S}} \leq 2\sqrt{2}\delta_j^{-1}||\hat{G}_{\mathcal{S}}(s,u) - G_{\mathcal{S}}(s,u)||, \tag{33}$$

where δ_j is defined in (A.5). By (A.5), Eq. (16) and Eq. (17) directly follow.

In the following, we establish Eq. (18) as follows,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \sup_{1\leq m\leq M_i} |\hat{\xi}_{ij}(t_{im}) - \xi_{ij}(t_{im})| \\
\leq &\frac{1}{n}\sum_{i=1}^n \sup_{1\leq m\leq M_i} |\int (X_i(s,t_{im}) - \mu(s,t_{im}))(\psi_j(s) - \hat{\psi}_j(s))ds| \\
&+ \frac{1}{n}\sum_{i=1}^n \sup_{1\leq m\leq M_i} |\int (\mu(s,t_{im}) - \hat{\mu}(s,t_{im}))\psi_j(s)ds| \\
&+ \frac{1}{n}\sum_{i=1}^n \sup_{1\leq m\leq M_i} |\int (\hat{\mu}(s,t_{im}) - \mu(s,t_{im}))(\psi_j(s) - \hat{\psi}_j(s))ds| \\
\leq &\frac{1}{n}\sum_{i=1}^n \sup_{s,t} |X_i(s,t_{im}) - \mu(s,t)| \|\psi_j(s) - \hat{\psi}_j(s)\|_{\mathcal{S}} \\
&+ \sup_{s,t} |\mu(s,t) - \hat{\mu}(s,t)| \sup_s |\psi_j(s)| + \sup_{s,t} |\hat{\mu}(s,t) - \mu(s,t)| \|\psi_j(s) - \hat{\psi}_j(s)\|_{\mathcal{S}} \\
&= O_p((\log n/n)^{1/2}) + O_p((\log n/n)^{1/2}) + O_p((\log n/n)^{1/2}) \\
&= O_p((\log n/n)^{1/2}),
\end{aligned} \tag{34}$$

where we used (A.1),(A.2), $\sup_{s,t}|\hat{\mu}(s,t)-\mu(s,t)| = O_p((\log n/n)^{1/2})$, and the previous result $\sup_s|\hat{\phi}_j(s)-\phi_j(s)| = O_p((\log n/n)^{1/2})$. This completes the proof.

Proof of Theorem 4:

$$\begin{aligned}
&C((s,t),(u,v)) = \mathrm{cov}(X(s,t),X(u,v))\\
=&\mathrm{cov}\left(\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\chi_{jk}f_k(t)g_j(s),\sum_{l=1}^{\infty}\sum_{h=1}^{\infty}\chi_{lh}f_h(v)g_l(u)\right)\\
=&\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{h=1}^{\infty}\mathrm{cov}(\chi_{jk}\chi_{lh})f_k(t)g_j(s)f_h(v)g_l(u).
\end{aligned}$$

Furthermore, by the orthogonality of f_k and $\mathrm{cov}(\chi_{jk},\chi_{lk}) = 0$ for $j \neq l$, we have

$$C_{\mathcal{S}}(s,u) = \int_{\mathcal{T}} C((s,t),(u,t))dt = \sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\sum_{h=1}^{\infty}\mathrm{cov}(\chi_{jk}\chi_{lh})g_j(s)g_l(u)\int_{\mathcal{T}} f_k(t)f_h(t)dt$$

$$=\sum_{j=1}^{\infty}(\sum_{k=1}^{\infty}\mathrm{cov}(\chi_{jk},\chi_{jk}))g_j(s)g_j(u).$$

Therefore, $g_j(s)$ are the unique eigenfunctions of $C_{\mathcal{S}}(s,u)$, and $\tau_j = \sum_{k=1}^{\infty}\mathrm{var}(\chi_{jk})$. By symmetry, one obtains the analogous result $f_k(t) \equiv \phi_k(t)$.

Proof of Theorem 5:

For $f_k(t)$ and $g_j(s)$ that satisfy the orthogonality conditions,

$$\begin{aligned}
&E\left(\int_{\mathcal{S},\mathcal{T}}\{X^c(s,t)-\sum_{j=1}^{P}\sum_{k=1}^{K}\langle X^c,f_kg_j\rangle f_k(t)g_j(s)\}^2dsdt\right)\\
=&E\|X^c\|^2 - 2\times\sum_{j=1}^{P}\sum_{k=1}^{K}\int E\left(X^c(s,t)X^c(u,v)\right)f_k(t)f_k(v)g_j(s)g_j(u)dsdtdudv\\
&+\sum_{j=1}^{P}\sum_{k=1}^{K}\int E\left(X^c(s,t)X^c(u,v)\right)f_k(t)f_k(v)g_j(s)g_j(u)dsdtdudv\\
=&E\|X^c\|^2 - \sum_{j=1}^{P}\sum_{k=1}^{K}\int E\left(X^c(s,t)X^c(u,v)\right)f_k(t)f_k(v)g_j(s)g_j(u)dsdtdudv.
\end{aligned} \tag{35}$$

Let $\tilde{f}_k(t)$ and $\tilde{g}_j(s)$ denote the optimal basis to achieve the minimum reconstruction error Q^*, and define

$$(I) = \sum_{j=1}^{P}\sum_{k=1}^{K}\int E\left(X^c(s,t)X^c(u,v)\right)\phi_k(t)\phi_k(v)\psi_j(s)\psi_j(u)dsdtdudv,$$

and

$$(II) = \sum_{j=1}^{P}\sum_{k=1}^{K}\int E\left(X^c(s,t)X^c(u,v)\right)\tilde{f}_k(t)\tilde{f}_k(v)\tilde{g}_j(s)\tilde{g}_j(u)dsdtdudv.$$

By Eq. (35), to prove the theorem, we only need to show that $(II) - (I) < aE\|X^c\|_2$.

We further define

$$(III) = \sum_{j=1}^{P}\int_{\mathcal{S}\times\mathcal{S}}\int_{\mathcal{T}} E\left(X^c(s,t)X^c(u,t)\right)dt\tilde{g}_j(s)\tilde{g}_j(u)dsdu,$$

and

$$(IV) = \sum_{j=1}^{P}\int_{\mathcal{S}\times\mathcal{S}}\int_{\mathcal{T}} E\left(X^c(s,t)X^c(u,t)\right)dt\psi_j(s)\psi_j(u)dsdu.$$

We will prove that $(II) < (III) < (IV)$ and $(IV) - (I) < aE\|X^c\|_2$, which implies that $(II) - (I) < aE\|X^c\|_2$.

By definition, the ψ_j are the leading eigenfunctions of the marginal kernel $G_{\mathcal{S}}(s,u)$ so that $(III) < (IV)$.

To show $(II) < (III)$, let $\tilde{\xi}_j(t) = \langle X^c(s,t), \tilde{g}_j(s)\rangle$, we observe

$$\begin{aligned}
(II) =& \sum_{j=1}^{P}\sum_{k=1}^{K}\int E\left(X^c(s,t)X^c(u,v)\right)\tilde{f}_k(t)\tilde{f}_k(v)\tilde{g}_j(s)\tilde{g}_j(u)dsdtdudv \\
=& \sum_{j=1}^{P}\sum_{k=1}^{K}\int E\left(\tilde{\xi}_j(t)\tilde{\xi}_j(v)\right)\tilde{f}_k(t)\tilde{f}_k(v)dtdv \\
<& \sum_{j=1}^{P}\int E\left(\tilde{\xi}_j(t)\tilde{\xi}_j(t)\right)dt \\
=& \sum_{j=1}^{P}\int E\left(\int X^c(s,t)\tilde{g}_j(s)ds\int X^c(u,t)\tilde{g}_j(u)du\right)dt \\
=& (III). \qquad (36)
\end{aligned}$$

Finally, we prove $(IV) - (I) < aE\|X^c\|_2 = \max(a_{\mathcal{S}}, a_{\mathcal{T}})E\|X^c\|_2$. Recall that τ_j and ϑ_k are the eigenvalues of $G_{\mathcal{T}}(s,u)$ and $G_{\mathcal{S}}(t,v)$. Then

$$\begin{aligned}
\tau_j &= \int G_{\mathcal{T}}(s,u)\psi_j(s)\psi_j(u)dsdu \\
&= \int\int E\left(X^c(s,t)X^c(u,t)\right)dt\psi_j(s)\psi_j(u)dsdu \\
&= \int\int \sum_{j'}\sum_{k}\sum_{l}\sum_{h} E(\chi_{j'k}\chi_{lh})\phi_k(t)\phi_h(t)dt\psi_{j'}(s)\psi_l(u)\psi_j(s)\psi_j(u)dsdu \\
&= \sum_{k=1}^{\infty}\mathrm{var}(\chi_{jk}),
\end{aligned}$$

and by symmetry, we obtain $\vartheta_k = \sum_{j=1}^{\infty}\mathrm{var}(\chi_{jk})$. Then,

$$\begin{aligned}
(IV) - (I) &= \sum_{j=1}^{P}\tau_j - \sum_{j=1}^{P}\sum_{k=1}^{K}\mathrm{var}(\chi_{jk}) = \sum_{j=1}^{P}\sum_{k=K+1}^{\infty}\mathrm{var}(\chi_{jk}) \\
&< \sum_{k=K+1}^{\infty}\vartheta_k = a_{\mathcal{T}}E\|X^c\|_2,
\end{aligned}$$

By symmetry, we also have $(IV) - (I) < a_{\mathcal{S}}E\|X^c\|_2$.

Supplement B: Additional Tables and Figures for the analysis of the fertility data and simulations

Additional materials on the fertility data, which were downloaded from the human fertility database on March 18, 2013, are provided in Table 4 and Figures 5-11. These complement the results presented in sections 5 and 6 of the main part of the paper.

Table 4: The abbreviations and names of the 17 countries (or territories), whose data are used in the fertility application (those with available data for the period 1951-2006). The colors used for representing each country in Figures 4 and 6 are also shown.

Color	Abbreviation	Country name	First year	Last year
	SWE	Sweden	1891	2010
	CAN	Canada	1921	2007
	ESP	Spain	1922	2006
	CHE	Switzerland	1932	2009
	USA	U.S.	1933	2010
	GBRTENW	U.K., England and Wales	1938	2009
	FIN	Finland	1939	2009
	PRT	Portugal	1940	2009
	GBR_SCO	U.K., Scotland	1945	2009
	FRA	France	1946	2010
	BGR	Bulgaria	1947	2009
	JPN	Japan	1947	2009
	CZE	Czech Republic	1950	2011
	HUN	Hungary	1950	2009
	NLD	Netherlands	1950	2009
	SVK	Slovakia	1950	2009
	AUT	Austria	1951	2010

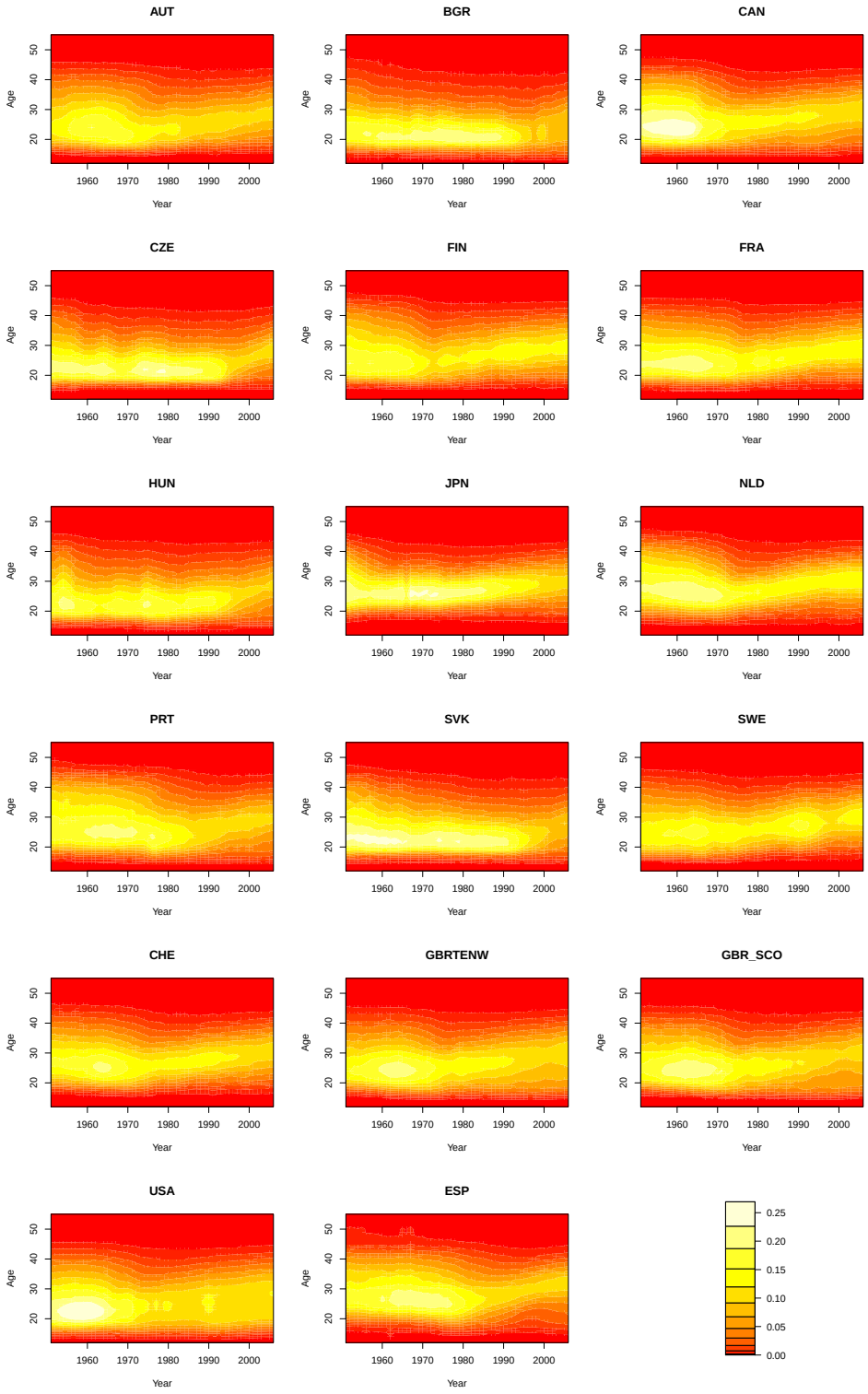

Figure 5: Age-specific fertility rates (ASFR) for 17 countries, red colors correspond to low values and yellow colors to high values.

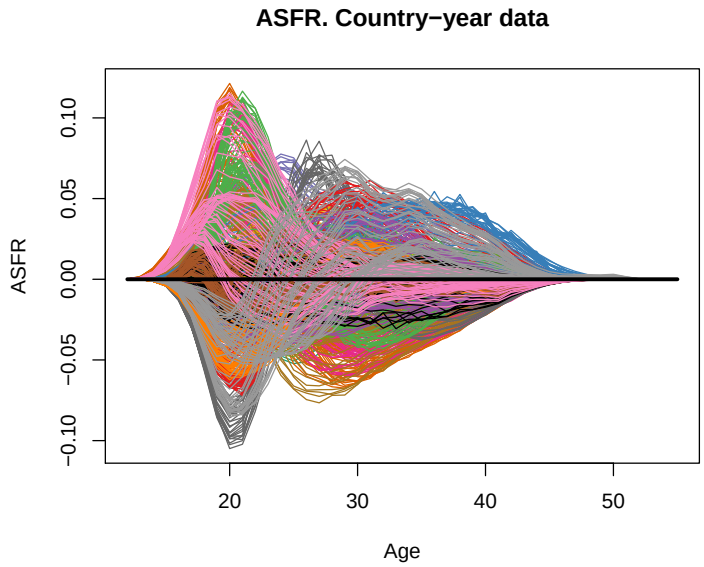

Figure 6: All available functional fertility data as functions of *age* for 952 combinations of 17 countries and 56 calendar years, centered around the mean. Functions corresponding to the same country are in the same color.

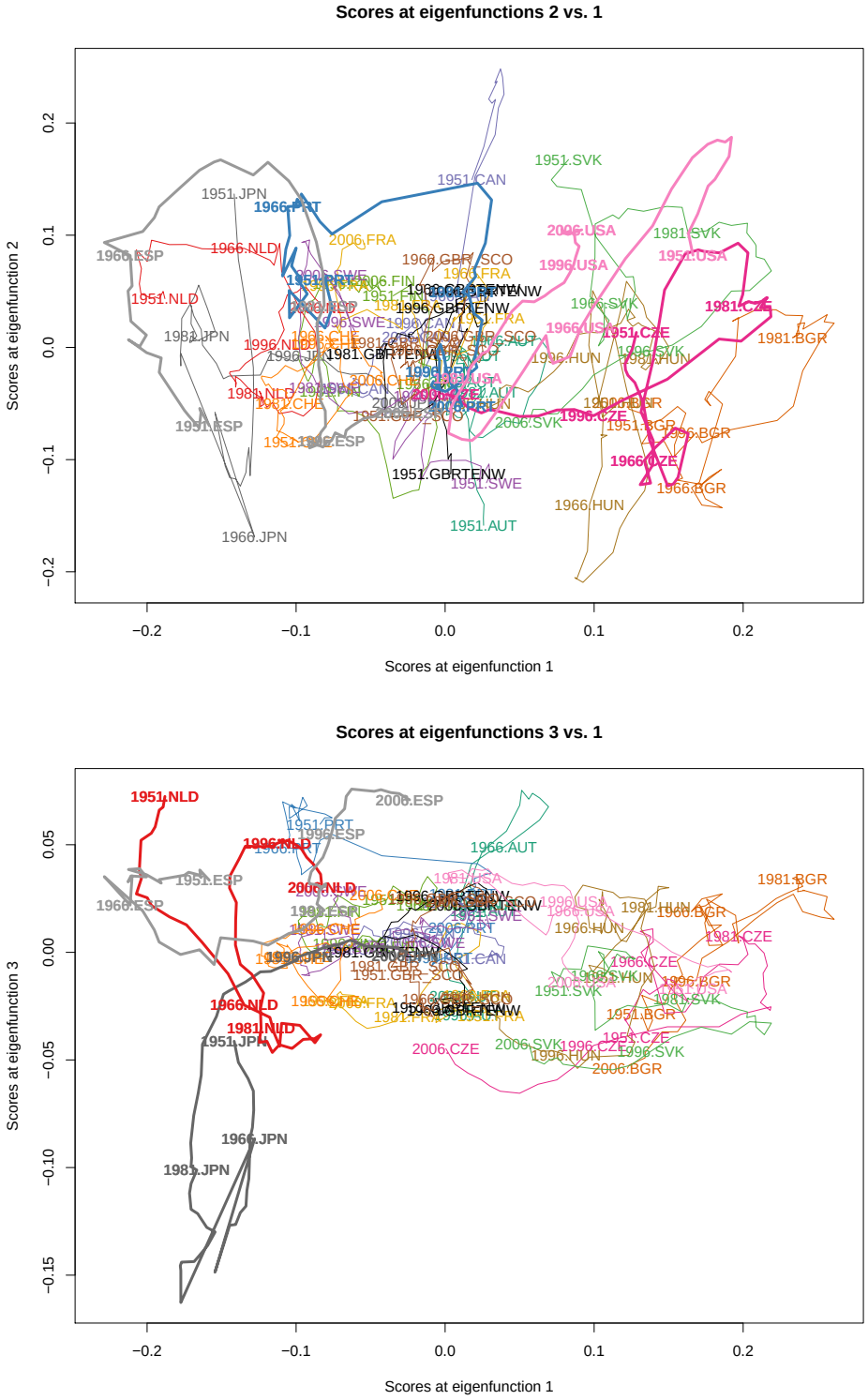

Figure 7: Track-plots corresponding to the implicitly parametrized planar curves $\{(\hat{\xi}_{i,1}(t), \hat{\xi}_{i,2}(t)),\ t = 1951, \ldots, 2006\}$, parametrized by calendar time t, where $\xi_{i,j}(t)$ is the j-th score function for country i as in (4).

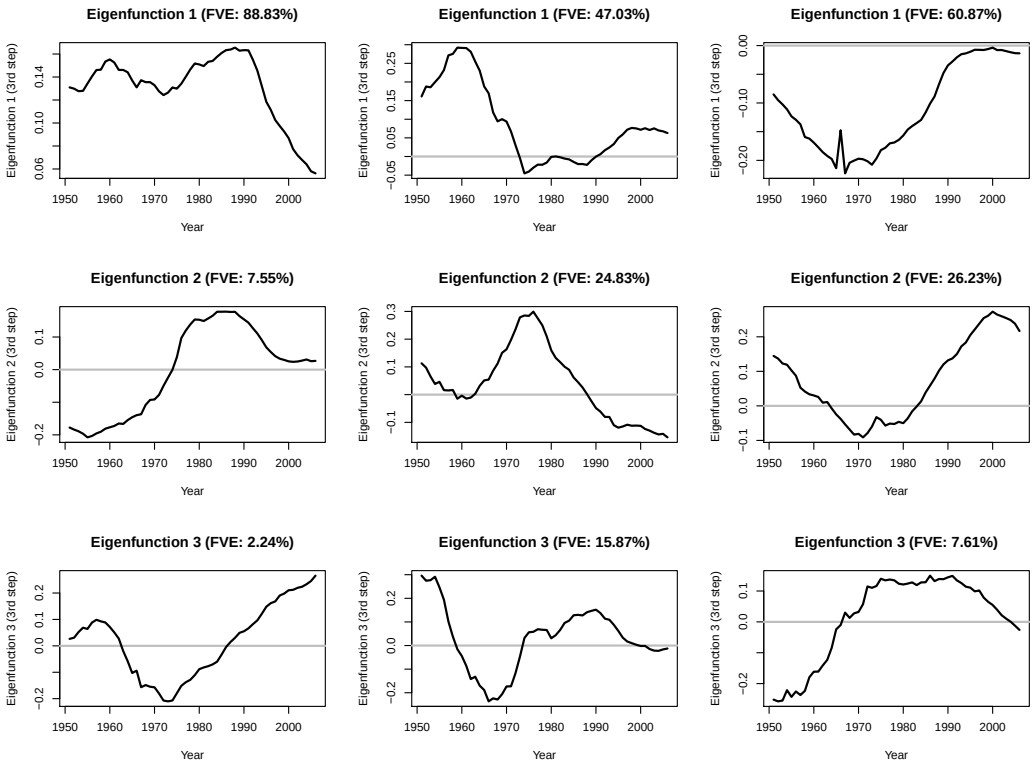

Figure 8: Estimated eigenfunctions $\hat{\phi}_{jk}(t)$ of the random scores $\xi_j(t)$. These quantities are as defined in (4).

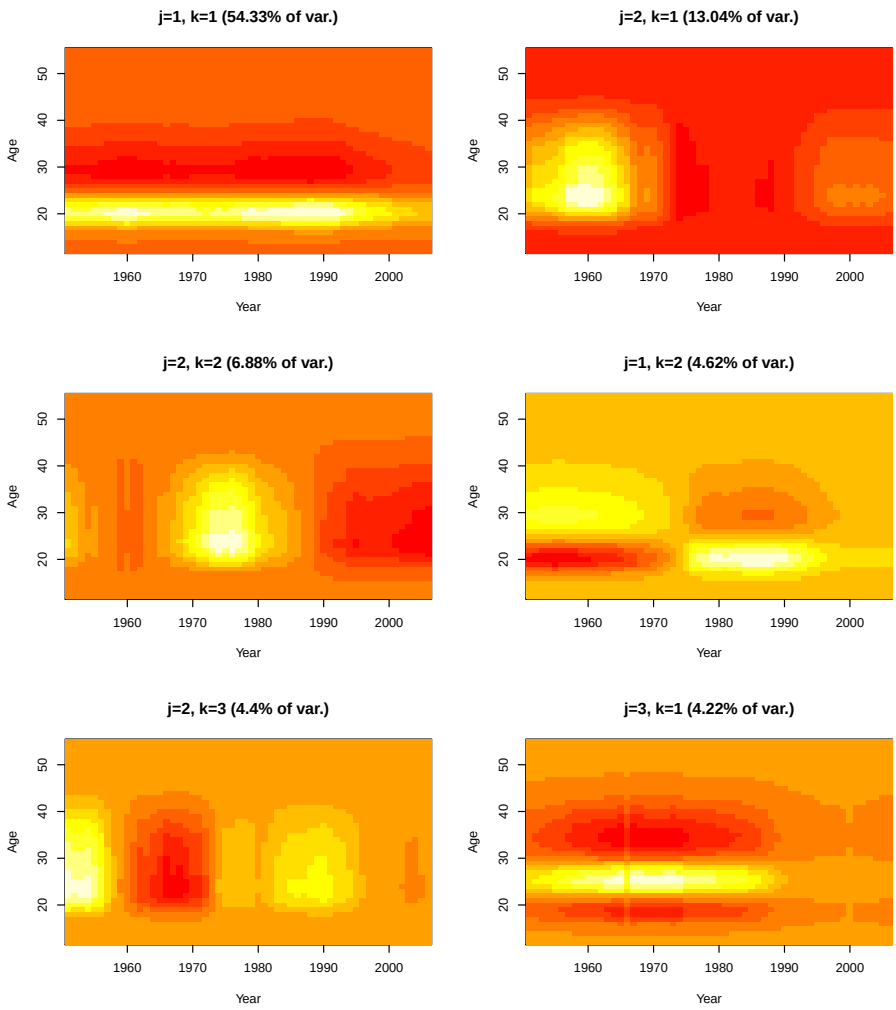

Figure 9: Product functions $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$ corresponding to the six terms with higher FVE in the marginal FPCA representation (4) of ASFR(s,t).

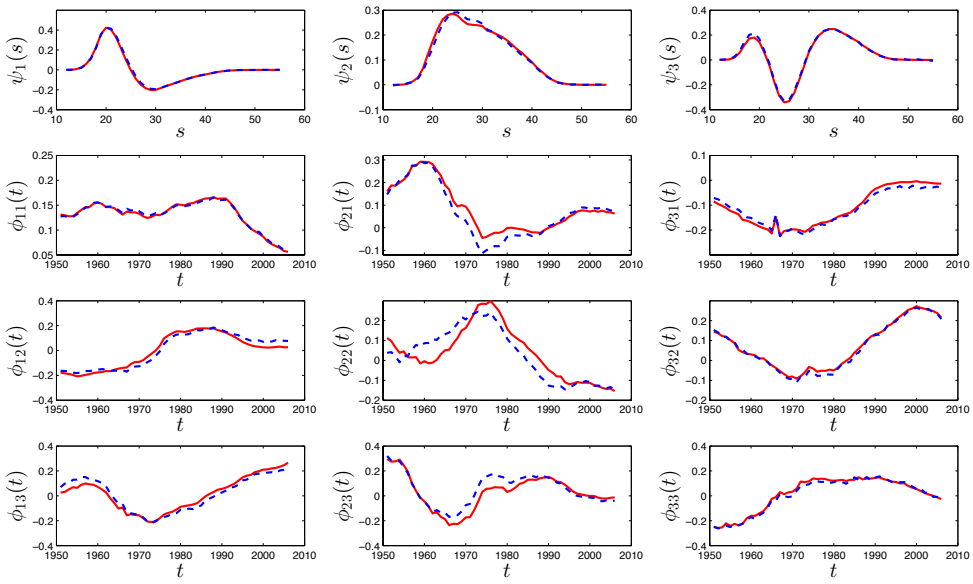

Figure 10: True (red-solid) and estimated (blue-dashed) eigenfunctions $\psi_j(s)$ and $\phi_{jk}(t)$ as in model (4) for $j = 1, 2, 3$ and $k = 1, 2, 3$, for one run of simulation 1 with sample size $n = 50$.

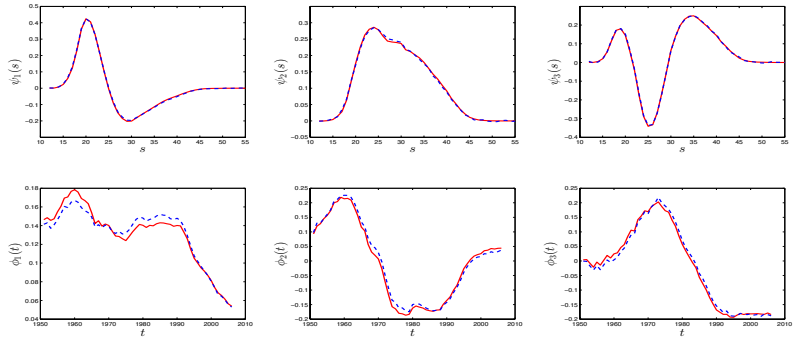

Figure 11: True (red-solid) and estimated (blue-dashed) eigenfunctions $\psi_j(s)$ and $\phi_k(t)$ in model (6) for $j = 1, 2, 3$ and $k = 1, 2, 3$, as obtained in one run of simulation 2 with sample size $n = 50$.

Supplement C: Standard two-dimensional FPCA and product FPCA for fertility data

Here we present the standard FPCA analysis of the ASFR data with the Karhunen-Loève representation, considering the data as random functions in two arguments. We performed FPCA for this type of functional data following Yao *et al.* (2005) as implemented in the PACE package (http://www.stat.ucdavis.edu/PACE). First, we rearrange the $n = 17$ matrices with dimension $L \times M = 44 \times 56$, containing the observed functional data, into a big data matrix with dimension $n \times (M \cdot L)$. Then we perform FPCA on this big matrix. Finally we rearrange the estimated eigenfunctions (stored at this point as arrays of length $M \cdot L$) into matrices of dimension $M \times L$. Figure 12 graphically summarizes the main results of this standard FPCA.

The first 4 eigenfunctions (which are eigensurfaces in this unconstrained approach) have a FVE of 89.73%. The first one (with FVE equal to 58.93%) is almost constant in calendar year and corresponds to a contrast between young fertility (women aged between 18 and 25 years) and fertility in mature years (mothers being from 25 to 40 years old). Countries with larger positive coefficients in this eigenfunction are Bulgaria, Czech Republic, Slovakia, Hungary and U.S., while the Netherlands, Japan, Spain and Switzerland have negative coefficients.

The second eigenfunction (or eigensurface) reflects the specificity of the baby-boom around 1960 in Canada and U.S. (both have high positive coefficients in this eigenfunction). Countries with negative scores (such as Japan, Spain, Bulgaria, Hungary or Czech Republic) do not show a drop in fertility rates at the end of the 1960s. The third eigenfunction appears to correspond to a sudden drop at the end of the 1970s in fertility for women aged between 30 and 40 years. This could be associated with women's decision on reducing the number of children, as the high fertility rates for ages in the interval [30,40] before 1977 are mainly associated with large families or, in more technical terms, with high parities, *parity* being defined as the cumulative number of a woman's live births; see Preston *et al.* (2001)). This

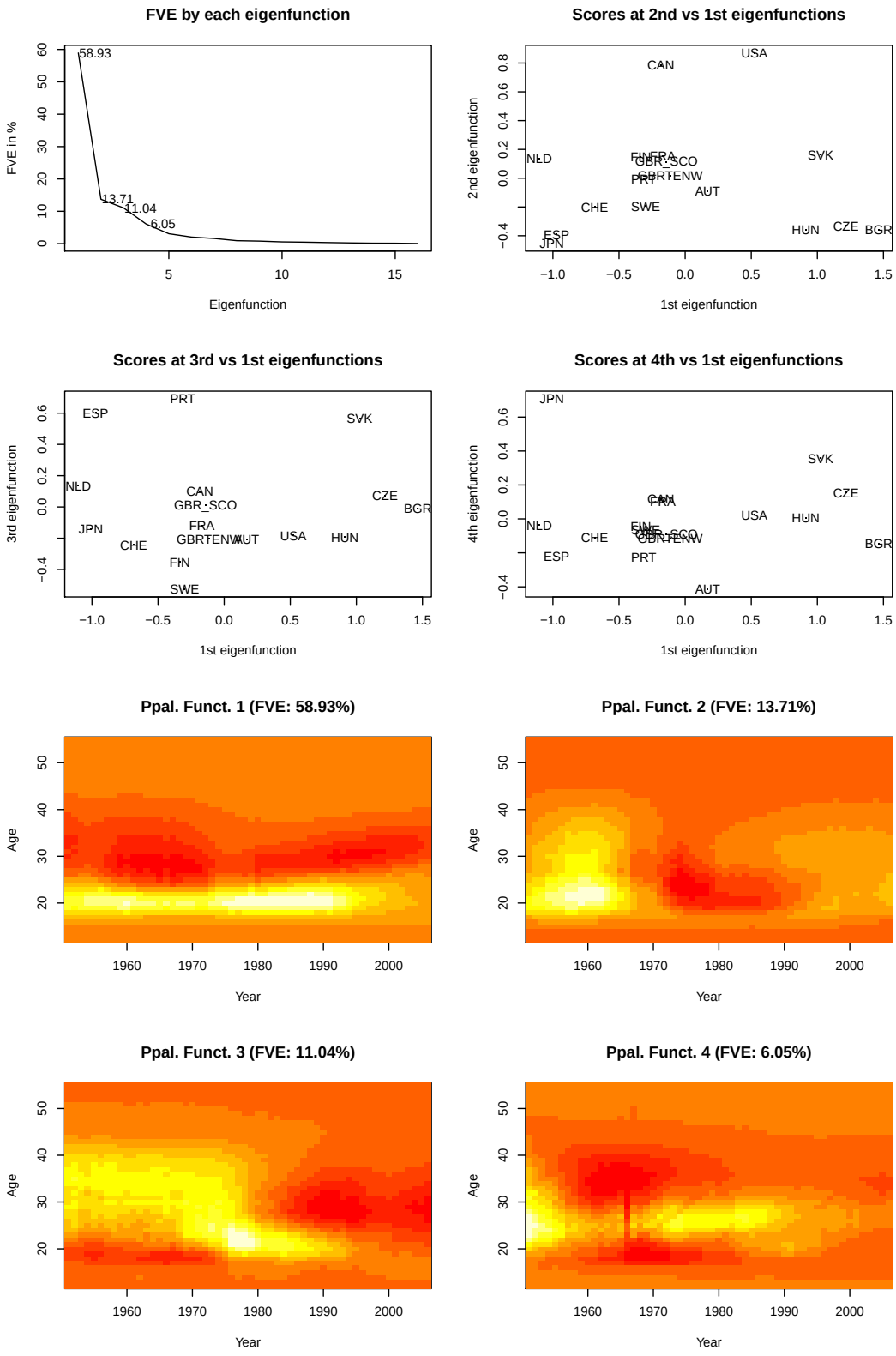

Figure 12: Standard FPCA of the fertility data $\text{ASFR}_i(s,t)$, $i = 1, \ldots, n = 17$, where the lower four panels display the first four eigensurfaces.

drop may be related to advances in birth control. These social changes arrived with a certain lag in countries with positive scores (Portugal, Spain, Slovakia) while they were adopted much earlier in countries with negative scores (Sweden, Finland, Switzerland). Other characteristics of this third eigenfunction are less intuitive.

Regarding the fourth eigenfunction, the score map in the panel in column 2, row 2, of Figure 12 indicates that Japan strongly weighs in this eigenfunction. Meanwhile the heat map (panel in column 2, row 4) shows a contrast between fertility concentrated around the age of 25 years (this strongly applies for Japan, with an outstanding positive score in this eigenfunction) and spread out fertility between the ages of 18 to 40, mainly between 1955 and 1980. Moreover, this heat map also shows an anomalous behavior (that appears as a discontinuity) at the year 1966. This fact corroborates that the fourth eigenfunction is a Japan specific function. We refer to the discussion in Section 5 for the anomaly in Japanese fertility in 1966.

Fitting the product FPC model to the fertility data resulted in estimates for the first four eigenfunctions ϕ_k of the operator $G_{\mathcal{T}}(t, v)$ as shown in Figure 13. The first of these time trend functions particularly weighs the pre-1990 fertility, while the others are contrasts between different calendar time periods. These estimated eigenfunctions are then multiplied with the age eigenfunction estimates $\hat{\psi}_j(s)$ of Figure 3 to obtain the product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$ that appear in the product FPC model representation (6) of ASFR(s, t). Figure 18 displays these product functions corresponding to the seven terms with larger FVEs among those with $j \leq 4$ and $k \leq 4$, which together explain 87.38% of the total variability; see also Table 1.

The product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$ in Figure 18 match well with the corresponding products $\hat{\phi}_{jk}(t)\hat{\psi}_j(s)$ in Figure 9 that result from the more general marginal approach (see Appendix B). These functions can thus be similarly interpreted as previously described in Section 5.1. The simplified product FPCA provides representations that are thus slightly less flexible and therefore explain somewhat less of the variance when compared with those

obtained from marginal FPCA, but have equally good, if not better, interpretability.

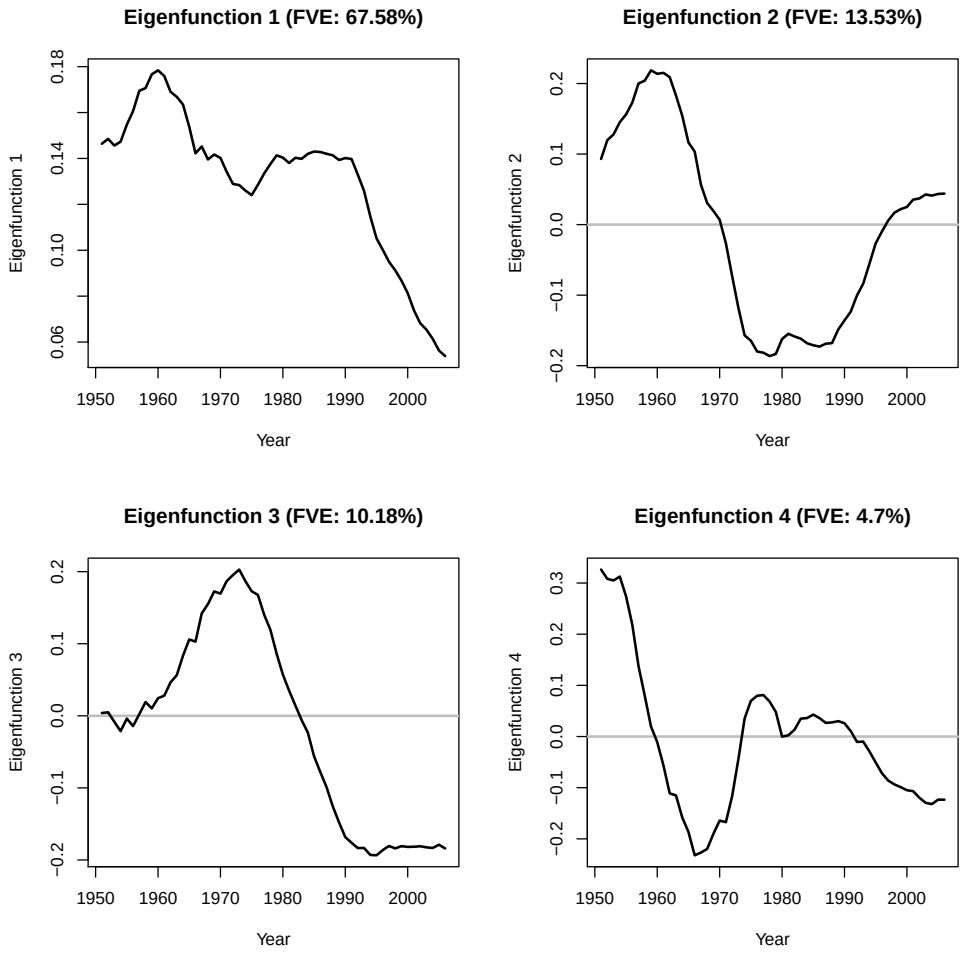

Figure 13: Estimated eigenfunctions $\hat{\phi}_k(t)$, $k = 1, 2, 3, 4$, in the product FPC model (6) for the fertility data.

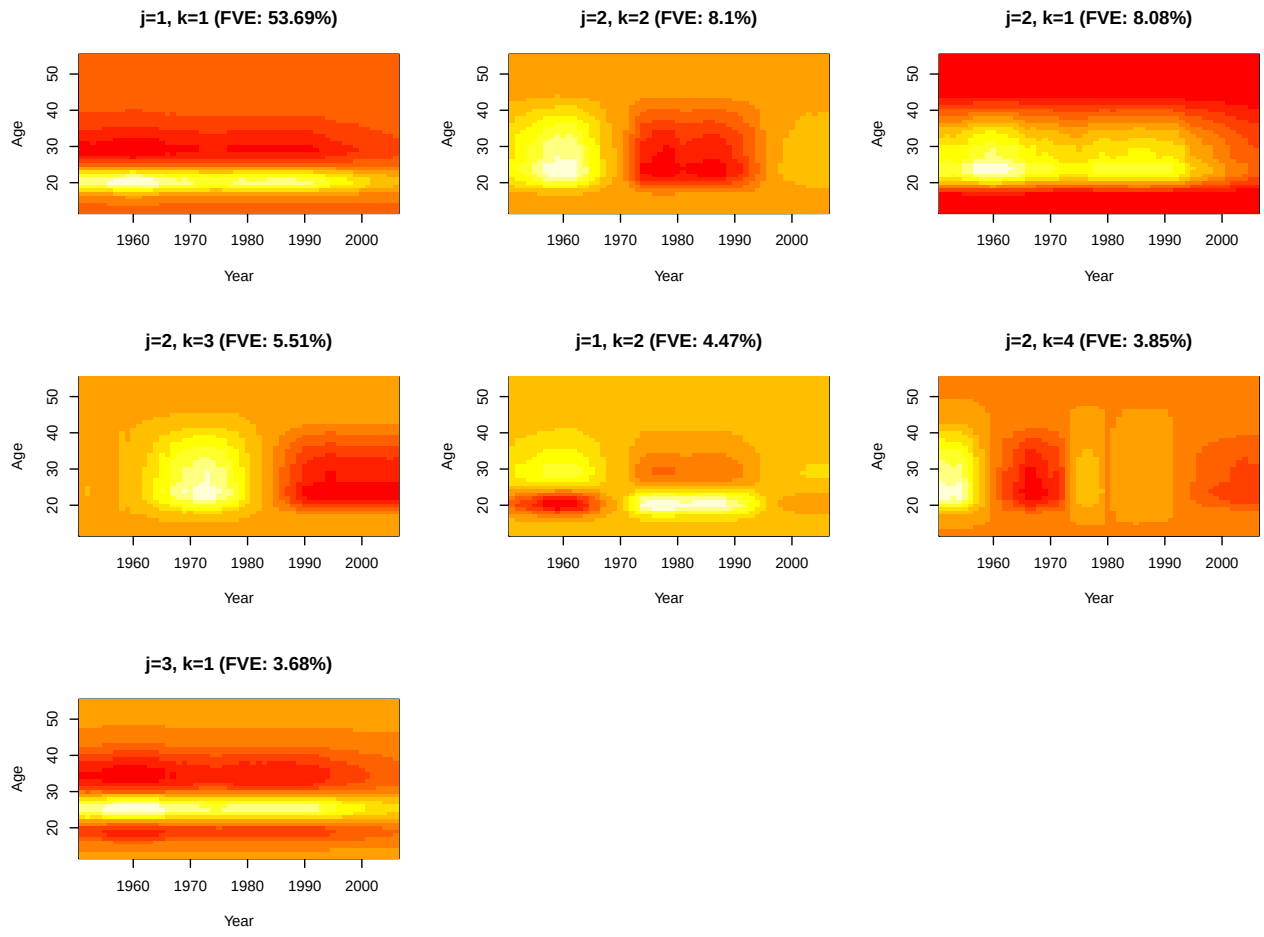

Figure 14: Product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$ corresponding to the seven terms with higher FVE in the product FPC model.

When applying product FPCA, one needs 7 terms to explain 87.38% of variance, while for the marginal FPCA it is sufficient to include 6 terms to explain 87.49% of the variance. Of course product FPCA has the big advantage that the final representation in general involves fewer functions $\hat{\psi}_j(s)$ and $\hat{\phi}_k(t)$ than the number of functions needed for the marginal FPCA representation and therefore is much simpler. For instance, the analysis of the fertility data with marginal FPCA involves 9 functions $(\hat{\psi}_1(s), \hat{\psi}_2(s), \hat{\psi}_3(s), \hat{\phi}_{11}(t), \hat{\phi}_{12}(t), \hat{\phi}_{21}(t), \hat{\phi}_{22}(t), \hat{\phi}_{23}(t), \hat{\phi}_{31}(t))$, while only 7 functions are involved in the product FPC model $(\hat{\psi}_1(s), \hat{\psi}_2(s), \hat{\psi}_3(s), \hat{\phi}_1(t), \hat{\phi}_2(t), \hat{\phi}_3(t), \hat{\phi}_4(t))$.

Supplement D: Male mortality rates as an additional example

Mortality rates (or death rates) are defined as a ratio of the death count for a given age-time interval divided by an estimate of the population exposed to the risk of death during some age-time interval (Preston *et al.* 2001). The Human Mortality Database (http://www.mortality.org/) provides detailed information on mortality rates for 37 countries or areas with precision of one year in both age and calendar time. Such rich information can be provided only by countries with well developed official statistics agencies. This is the reason why only 37 countries are covered by this database.

An alternative database including a much larger number of countries can be accessed through the Population Division of the Department of Economic and Social Affairs of the United Nations (WPP 2012). This database contains information for more than 200 countries on deaths grouped into five-year age intervals, from 1950 to 2010 (every 5 years). The price to be paid for including a much larger number of countries is a loss in precision, i.e., aggregation over 5 year intervals, both in terms of age and calendar time. As definition of the mortality rate for a given country during a period of consecutive years and an interval of ages, we consider the ratio between the number of deaths reported for a specific country over the selected 5 year calendar period for people with age at death in the selected 5 year age interval, divided by the number of people that at the beginning of the calendar time interval were in the age interval. As male and female mortality rates are different, we consider here male data that were downloaded (on the 14th of January 2015) from

```
http://esa.un.org/wpp/Excel-Data/EXCEL_FILES/3_Mortality/
                        WPP2012_MORT_F04_2_DEATHS_BY_AGE_MALE.XLS
http://esa.un.org/wpp/Excel-Data/EXCEL_FILES/1_Population/
                        WPP2012_POP_F15_2_ANNUAL_POPULATION_BY_AGE_MALE.XLS
```

We work with *log-Mortality Rates*, for which we use log(mortality rate+1), considered as functions of men's age grouped into intervals of 5 years (s) and repeatedly measured for

every 5 calendar years t for various countries. The aggregated log-mortalities constitute the functional data $X(s,t) = \text{log-mortality rate}(s,t)$.

In WPP (2012), data are provided for ages s in the year intervals $\{[0,5), [5,10), \ldots, [90,95), [95,\infty)\}$. The interval of calendar years with available data are $\{[1950,1955), \ldots, [2005,2010)\}$. The variability of mortality rates increases with age and decreases with population. So we limit ourselves to ages lower than 80. We also excluded countries with a 0 value for population size at any year or age. Then our database finally consisted of 166 countries, with 12 periods of five years each (which we labeled with the first year of the respective interval: 1950 to 2005) and 16 five years age intervals (labeled from 0 to 75).

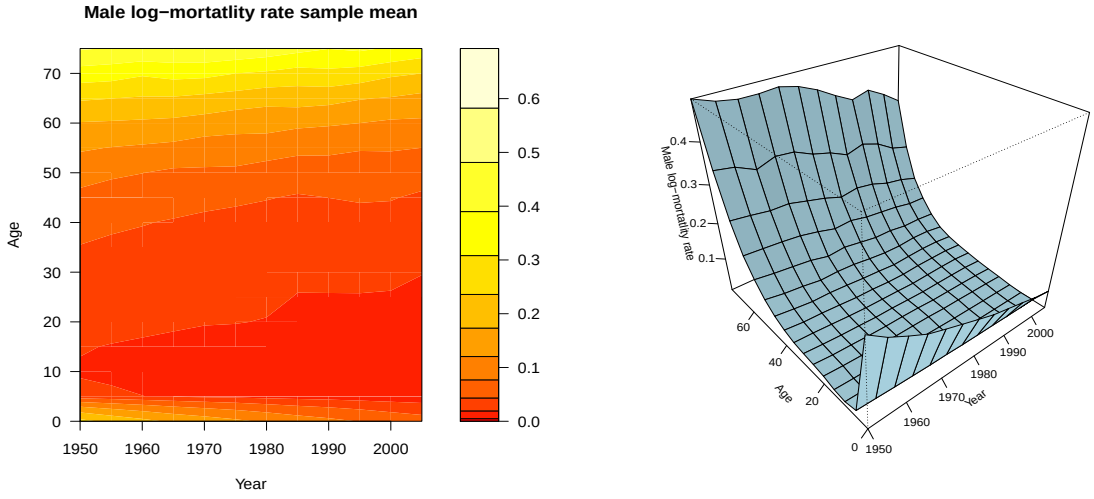

Figure 15: Sample mean of the 166 male log-mortality rate functions by calendar year.

The sample mean of the male log-mortality rate functions for 166 countries displayed in Figure 15 shows that mortality rates are, on average, highest for children under 5, and for men aged more than 60; and that they are decreasing with increasing calendar year.

The male log-mortality rate data include one log-mortality rate curve over age per calendar year and per country and are observed on a regular grid spaced in years across both coordinates age s and calendar year t, which means that the empirical estimators described in Section 2 can be applied to these data. Figure 16 displays the $nM = 1992$ centered functional data male log-mortality rates $X_i^c(s_l, t_m) = X_i(s_l, t_m)$ for $l = 1, \ldots, L = 16$,

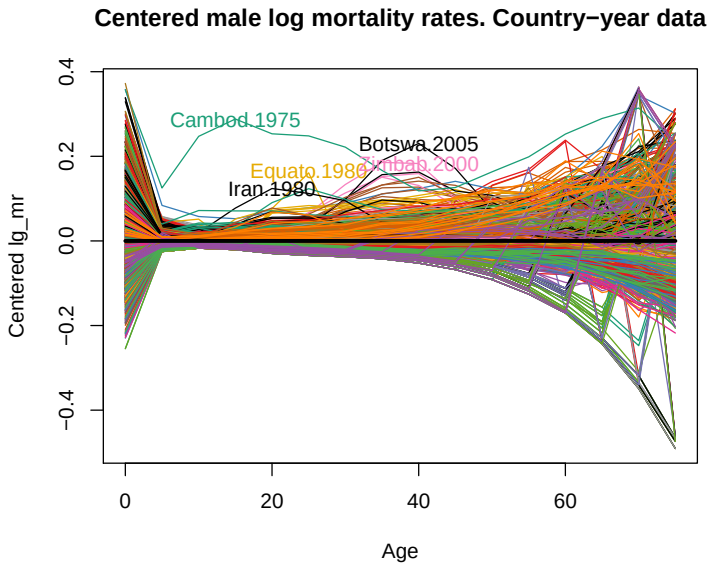

Figure 16: All available functional male log-mortality data as functions of *age* for 1992 combinations of 166 countries and 12 calendar years, centered around the mean. Functions corresponding to the same country are in the same color.

$m = 1, \ldots, M = 12$ and $i = 1, \ldots, n = 166$, demonstrating that there is substantial variation across countries and calendar years. Several outliers in the centered log mortality profiles have been highlighted in the figure. Some of these reflect periods of war, e.g. Iran 1980-1985 or genocides, e.g. Cambodia 1975-1980. Others correspond to high mortality rates due to the HIV/AIDS pandemic. The bloody reign of the Macias Nguema dictatorship in Equatorial Guinea also left its mark in this country's mortality profile.

We fitted the marginal FPC model and found that the $\phi_{jk}(t)$ are similar for $j = 1$ and 2. This is an indication that the product FPCA is appropriate for these data, and we directly applied it. Fitting the product FPC model to the male log-mortality data resulted in estimates for ψ_j and ϕ_k as shown in Figure 17. The shape of the first eigenfunction $\hat{\psi}_1(s)$ (that takes positive values for all ages s) is similar to that of the mean function for a fixed year t (see the right panel of Figure 15). Therefore $\hat{\psi}_1(s)$ can be interpreted as a *size* component: Country-years with positive score in the direction of this eigenfunction have higher male log-mortality ratios than the mean function for all ages, with larger differences

for larger values of the average log-mortality rates. The second eigenfunction $\hat{\psi}_2(s)$ represents a contrast between infant mortality and older age mortality. The third eigenfunction $\hat{\psi}_3(s)$ appears to point to difficulties in obtaining accurate estimates of mortality rates for the last age interval.

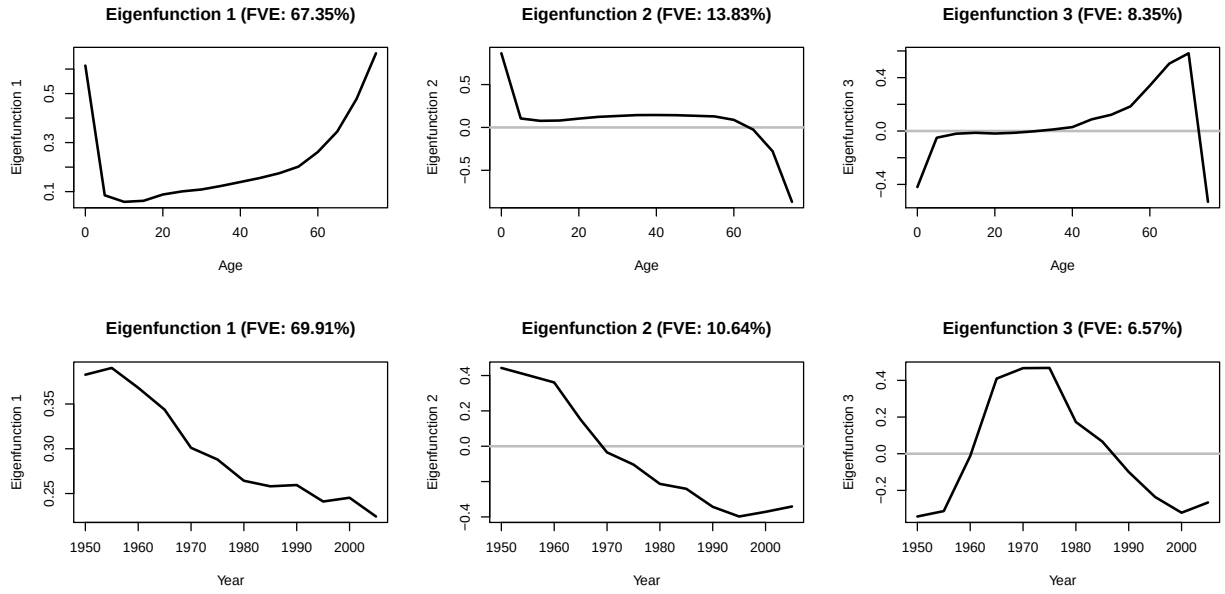

Figure 17: Estimated eigenfunctions $\hat{\psi}_j(s)$ (first row) and $\hat{\phi}_k(t)$ (second row), in the product FPC model for the log-mortality data.

The first calendar year trend function $\hat{\phi}_1(t)$ shows a continuous reduction in male log-mortality rates, with a pattern similar to the average evolution of male log-mortality over time (see Figure 15, right panel). So positive scores associated with this eigenfunction indicate larger reductions than the average (the opposite for negative scores). The second and third trend functions are contrasts between different calendar time periods. Positive (resp., negative) scores in the second trend function $\phi_2(s)$ indicate higher (resp., lower) than average mortality at the beginning of the overall calendar period, and lower than average mortality for the final years of the calendar period, i.e., a faster decline in mortality as compared to the average decline. The third eigenfunction is associated with differences in changes in log-mortality rates over calendar time between the middle period and the early/late periods.

The product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$ are shown in Figure 18. These functions can be easily interpreted by taking into account that they are the product of a function $\hat{\psi}_j(s)$ and a function $\hat{\phi}_k(t)$, as represented in Figure 17. When applying product FPCA, one needs 4 terms to achieve a FVE of 71.06%, and 6 terms to achieve a FVE of 75%.

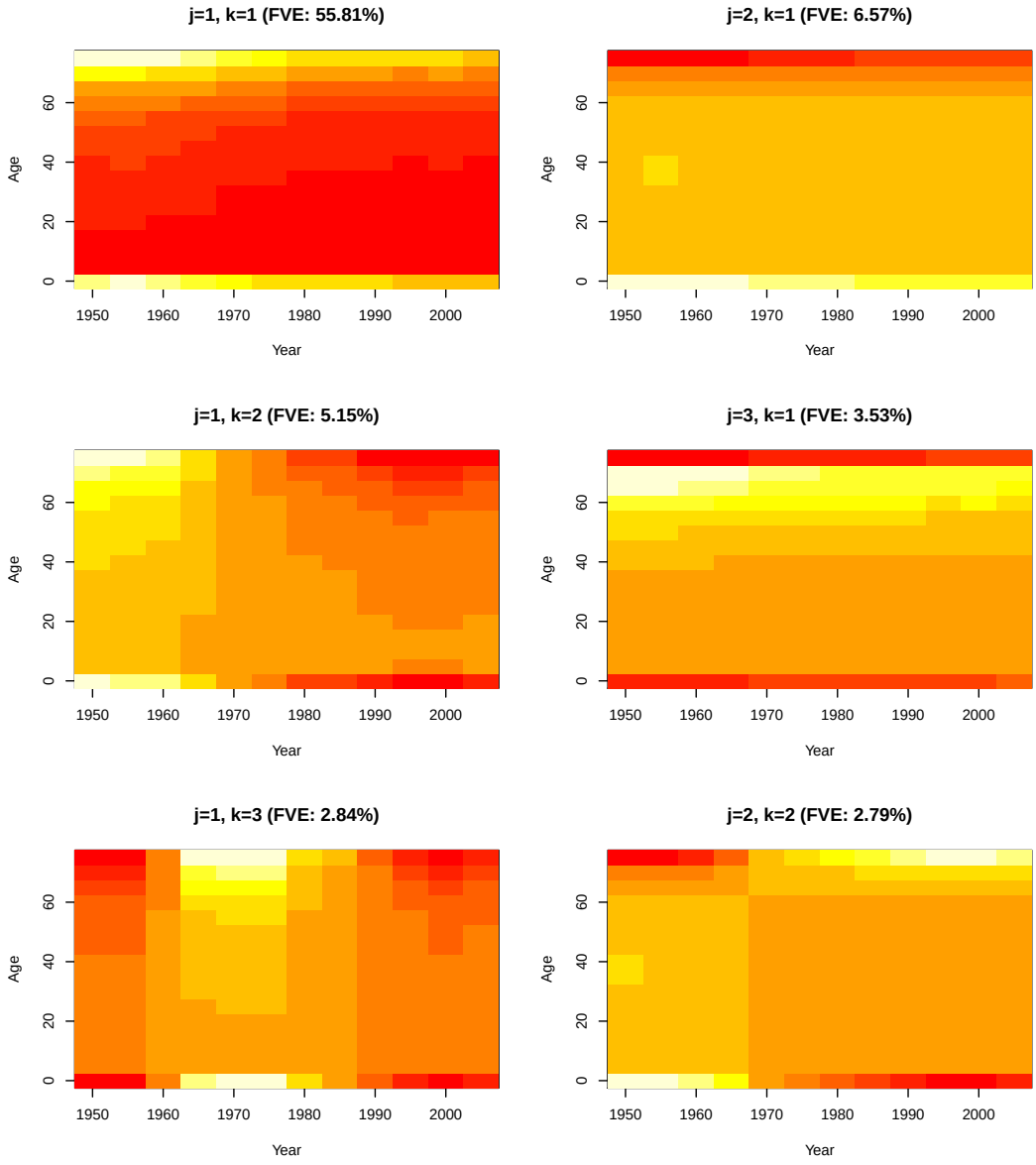

Figure 18: Product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$ corresponding to the six terms with higher FVE in the product FPC model representation for the log mortality data.

The first product of estimated eigenfunctions (with FVE equal to 55.81%) is $\hat{\phi}_1(t)\hat{\psi}_1(s)$, which is the product of the function $\hat{\phi}_1(t)$ that is similar to the average evolution of log-mortality rate over calendar years, and the function $\hat{\psi}_1(s)$ that has a shape similar to the average log-mortality rate pattern. As a consequence, the product function is always positive and very similar to the mean function (see Figure 15). So countries with positive random coefficients $\hat{\chi}_{11}$ at this product function $\hat{\phi}_1(t)\hat{\psi}_1(s)$ have larger male log-mortality rates than the average for all ages and all calendar years, with larger differences for larger values of the average log-mortality rates, and vice versa for the countries with a negative coefficient. We refer to Table 5 for a list of countries with most extreme (positive or negative) coefficients at this first product component.

The second product of eigenfunctions is $\hat{\phi}_1(t)\hat{\psi}_2(s)$ (FVE: 6.57%). It represents a contrast between infant mortality and old age mortality, due to the shape of $\hat{\psi}_2(s)$, which is more marked at the beginning than at the end of the period (because of $\hat{\phi}_1(t)$). Countries with negative scores (see Table 5) have lower than average infant log-mortality rates and higher than average old age log-mortality rates. The opposite applies to countries with positive scores at this product. The third product of eigenfunctions is $\hat{\phi}_2(t)\hat{\psi}_1(s)$ and it separates countries with faster than average reduction in male log-mortality rates (positive coefficients) from those with slower than average reduction (negative coefficients). This is the main effect of $\hat{\phi}_2(t)$. This effect is more marked for extreme ages, due to the shape of $\hat{\psi}_1(s)$. The countries with extreme coefficients as listed in Table 5 are extremes in a certain shape direction and deserve further study.

Alternatively, one can also apply marginal FPCA to quantify the observed variability across countries. The results for marginal FPCA are summarized in Figures 19 and 20 for the first three eigenfunctions, $\hat{\psi}_j(s)$, $j = 1, 2, 3$, resulting in a FVE of 89.52%. The first row of Figure 19 displays the estimated eigenfunctions $\psi_j(s)$, which are identical to the functions $\psi_j(s)$ used in the product FPCA. The second row of panels in Figure 19 shows the score functions $\hat{\xi}_{i,j}(t)$, $t \in \mathcal{T}$, which are country-specific functions of calendar year.

Table 5: Countries with the most extreme estimates of the random coefficients χ_{jk} obtained by fitting the product FPCA model (6) for the six terms with higher FVE in the representation of male log-mortality rates as linear combinations of the product functions $\hat{\phi}_k(t)\hat{\psi}_j(s)$.

	$\hat{\phi}_1(t)\hat{\psi}_1(s)$ (FVE: 55.81%)
Most −	Iceland, Channel Islands, Sweden, Norway, Puerto Rico, Barbados
Most +	Sierra Leone, Mali, Eritrea, Equatorial Guinea, Timor-Leste, Liberia
	$\hat{\phi}_1(t)\hat{\psi}_2(s)$ (FVE: 6.57%)
Most −	Fiji, Suriname, Martinique, Mauritius, Dem People's Republic of Korea, Guyana
Most +	Reunion, Central African Republic, El Salvador, Honduras, Pakistan, Angola
	$\hat{\phi}_2(t)\hat{\psi}_1(s)$ (FVE: 5.15%)
Most −	Channel Islands, Barbados, Iceland, Belarus, Rwanda, Sierra Leone
Most +	China, Oman, Tunisia, Singapore, Hong Kong SAR, Japan
	$\hat{\phi}_1(t)\hat{\psi}_3(s)$ (FVE: 3.53%)
Most −	Channel Islands, Iceland, Martinique, Guinea-Bissau, Timor-Leste, Oman
Most +	Reunion, Papua New Guinea, Eritrea, South Africa, Dem People's Republic of Korea, Guadeloupe
	$\hat{\phi}_3(t)\hat{\psi}_1(s)$ (FVE: 2.84%)
Most −	Cape Verde, Tajikistan, Kazakhstan, Azerbaijan, Belarus, Kyrgyzstan
Most +	Cambodia, Barbados, Channel Islands, Reunion, Guadeloupe, Martinique
	$\hat{\phi}_2(t)\hat{\psi}_2(s)$ (FVE: 2.79%)
Most −	Martinique, Japan, Fiji, Malta, Guyana, Botswana
Most +	Reunion, Barbados, Channel Islands, Iceland, Yemen, Eritrea

Their evolution over calendar time can be visualized by the track plot in Figure 20, showing the planar curves for the pairs $(\xi_{i,1}(t), \xi_{i,2}(t))$, $t \in \mathcal{T}$. In this example, the track plot is particularly useful to detect country-years with extreme scores in some eigenfunctions. For instance, Cambodia 1975-1980 and Rwanda 1990-1995 have extremely positive high scores in the first eigenfunction. This corresponds to periods in the history of these two countries during which they experienced a very high mortality rate: the Cambodian Genocide from 1975 to 1979, and the Rwandan Genocide in 1994.

The third step of the marginal FPCA (performing a separate standard FPCA for the estimated score functions $\hat{\xi}_{i,j}(t)$, $i = 1, \ldots, n$, for $j = 1, 2, 3$) yields estimated eigenfunctions $\hat{\phi}_{jk}$. For $k = 1, 2, 3$ these estimates are shown in Figure 19 (three lower rows). It can be seen that results are similar (up to sign changes) for the first and second sets of score functions.

To conclude this second example, we present the standard FPCA of the log-mortality data with the Karhunen-Loève representation, considering the data as random functions in two arguments. Figure 21 graphically summarizes the main results of this standard FPCA. The first four eigenfunctions have a FVE of 78.55%. There are similarities between these eigenfunctions and, respectively, the 1st, 2nd, 3rd and 5th eigenfunction products represented in Figure 18 (in the two last cases, up to a sign change). Therefore the interpretation we have made above for these eigenfunctions products are valid for the eigenfunctions obtained by standard FPCA. Nevertheless, to arrive at these interpretations is much more difficult if the starting point is Figure 21, without the benefit of the functions represented in Figure 18 for the product FPCA and their decomposition as products of functions in Figure 17.

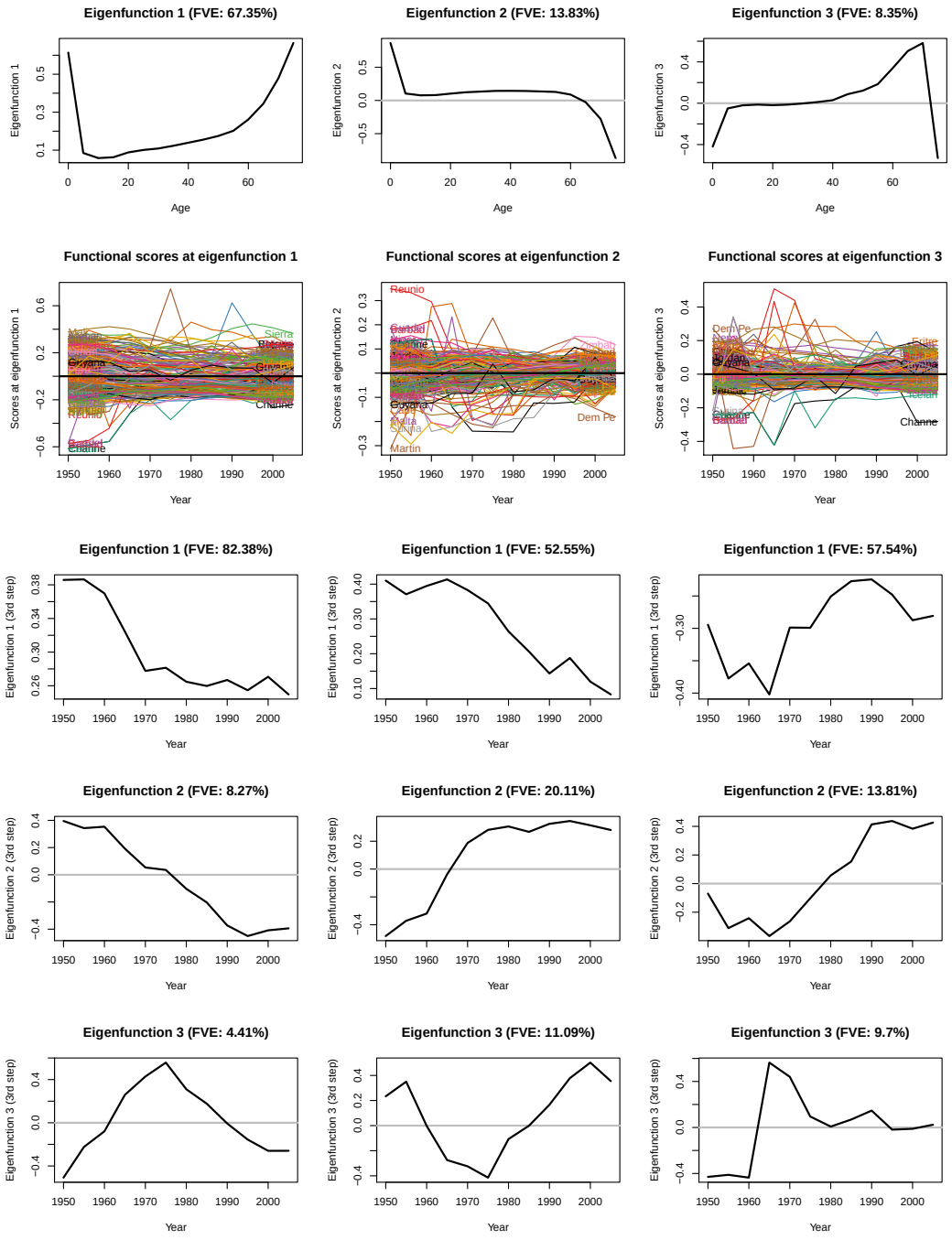

Figure 19: Results of the marginal FPCA for the male log-mortality rate data. Columns 1, 2 and 3 correspond to eigenfunctions 1, 2 and 3 in the second step of the marginal FPCA, respectively. *First row:* Estimated eigenfunctions $\hat{\psi}_j(s)$, where s is age. *Second row:* Score functions $\hat{\xi}_{i,j}(t)$, where t is calendar year. *Rows 3, 4 and 5:* Estimated eigenfunctions $\hat{\phi}_{jk}(t)$, $k = 1, 2, 3$, in the third step.

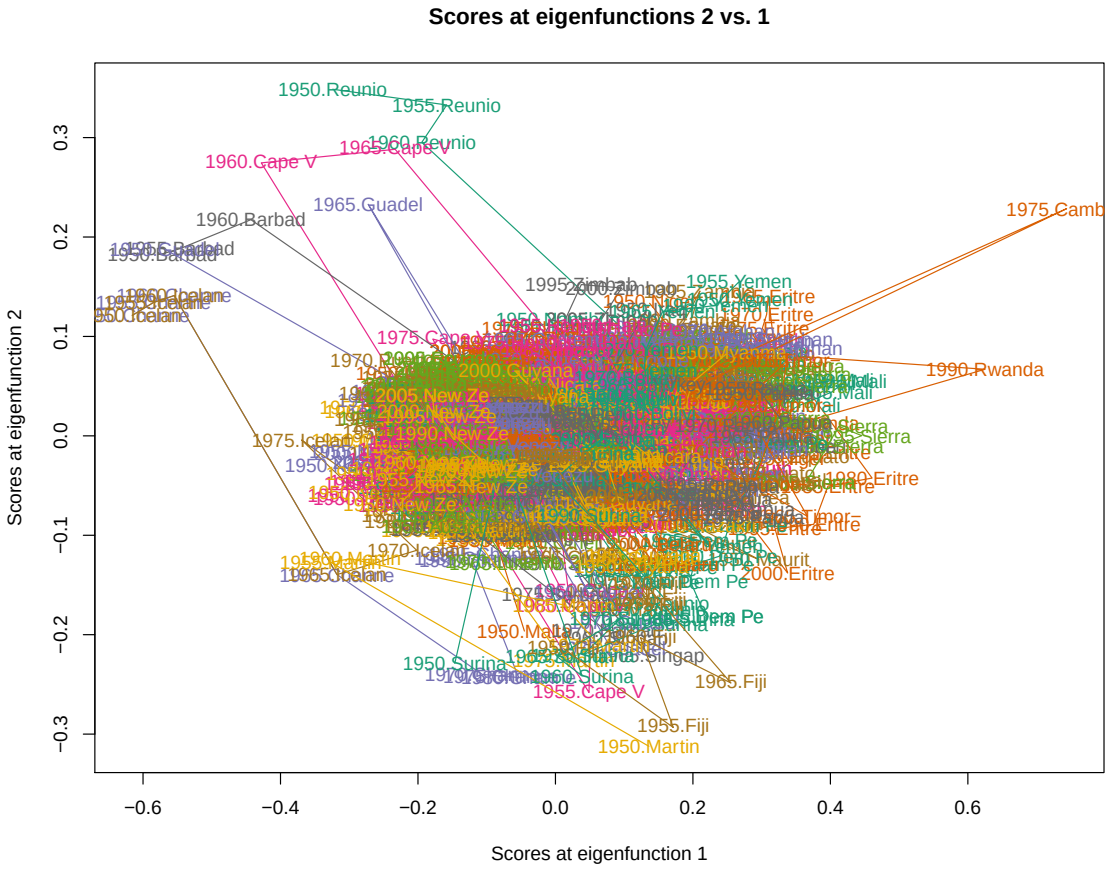

Figure 20: Track-plot corresponding to the implicitly parametrized planar curves $\{(\hat{\xi}_{i,1}(t), \hat{\xi}_{i,2}(t)) : t = 1950, 1955, \ldots, 2005\}$, parametrized by calendar time t, where $\xi_{i,j}(t)$ is the j-th score function for country i.

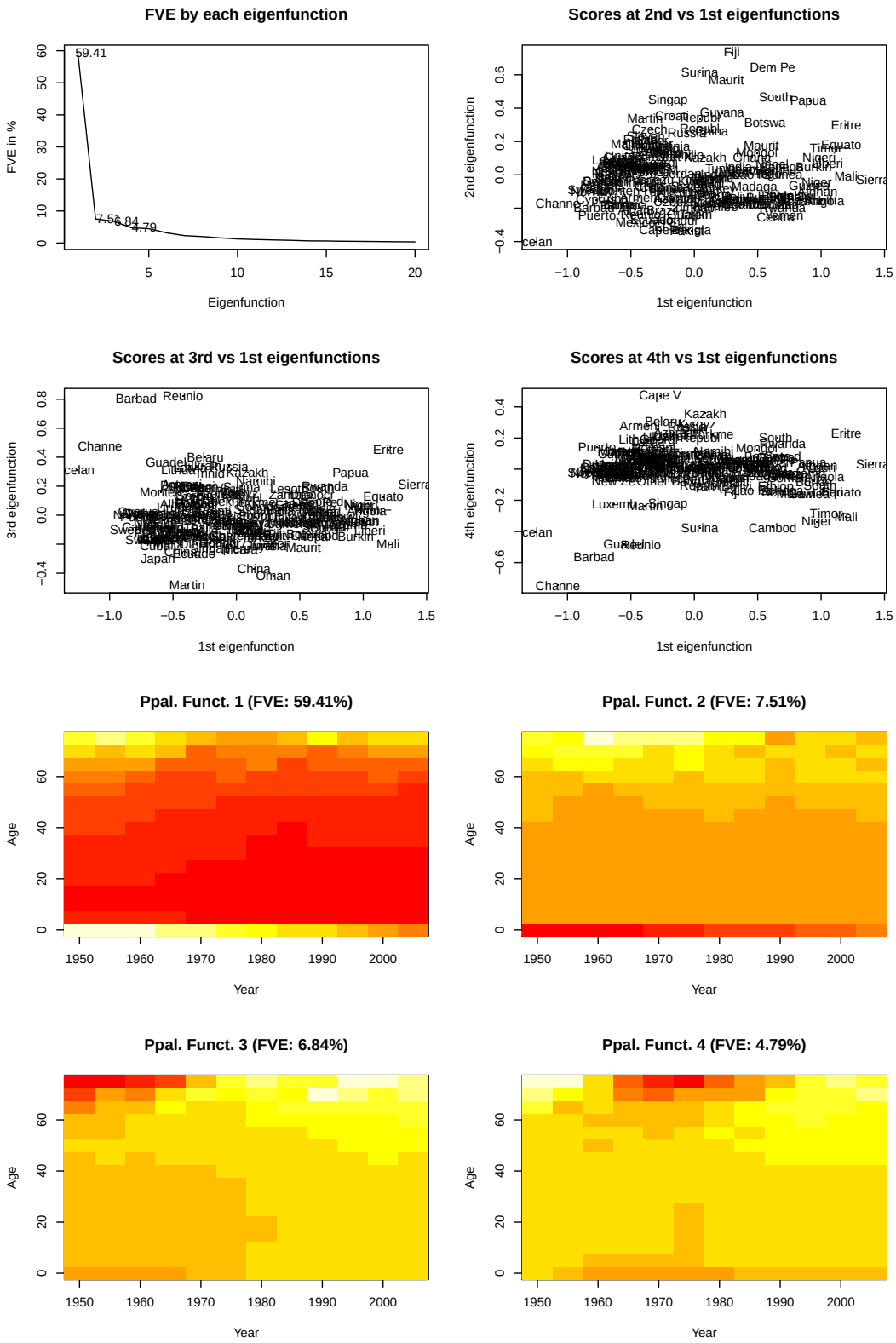

Figure 21: Standard FPCA of the male log-mortality data.